NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

VERIFICATION OF THE AFWA 3-ELEMENT SEVERE WEATHER FORECAST ALGORITHM

by

Daniel E. Pagliaro

March 2008

Thesis Advisor:	Wendell A. Nuss
Second Reader:	David S. Brown

Approved for public release; distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** March 2008	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE Verification of the AFWA 3-Element Severe Weather Forecast Algorithm			**5. FUNDING NUMBERS**
6. AUTHOR(S) Daniel E. Pagliaro			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Weather Agency 106 Peacekeeper Drive, Suite 2N3 Offutt AFB, NE 68113-4039			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Accurate severe thunderstorm forecasts are critical to providing sufficient lead-time to protect lives and property. The Air Force Weather Agency has developed a 3-Element Severe Weather Forecast Algorithm that when applied to model forecasts gives and outlook region for severe thunderstorms. Improvements were made in this study to enhance the algorithm's forecast skill, reduce its "false alarm" rate, and thereby increase the amount of lead-time for installation commanders to take decisive action to protect personnel and resources. This paper discusses the performance of the 3-Element Algorithm in its original form, and the adjustments made to overcome some of its limitations.

The 3-Element Algorithm techniques and results of a performance evaluation are presented. Based on the amount of forecast improvement, eight configurations were retained for analysis across the entire dataset containing six severe weather cases. A new stability proxy, the Elevated Total-Totals Index, was developed and integrated into the algorithm to improve severe weather forecasts over high-elevation regions where some traditional severe weather indices cannot be accurately computed. Additionally, the horizontal gradient of convective available potential energy was studied as a new indicator to the presence of dynamic forcing. It is hoped that improvements discussed in this paper will make the 3-Element Algorithm an effective tool in the early forecasting of severe weather, increasing lead-time to safeguard lives and resources.

14. SUBJECT TERMS Numerical Weather Prediction; Severe Local Storms			**15. NUMBER OF PAGES** 107
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited.

VERIIFCATION OF THE AFWA 3-ELEMENT SEVERE WEATHER FORECAST ALGORITHM

Daniel E. Pagliaro
Captain, United States Air Force
B.A., Western Connecticut State University, 2002

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY

from the

NAVAL POSTGRADUATE SCHOOL
March 2008

Author: Daniel E. Pagliaro

Approved by: Wendell A. Nuss
Thesis Advisor

David S. Brown
Second Reader

Philip A. Durkee
Chairman, Department of Meteorology

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Accurate severe thunderstorm forecasts are critical to providing sufficient lead-time to protect lives and property. The Air Force Weather Agency has developed a 3-Element Severe Weather Forecast Algorithm that when applied to model forecasts gives and outlook region for severe thunderstorms. Improvements were made in this study to enhance the algorithm's forecast skill, reduce its "false alarm" rate, and thereby increase the amount of lead-time for installation commanders to take decisive action to protect personnel and resources. This paper discusses the performance of the 3-Element Algorithm in its original form, and the adjustments made to overcome some of its limitations.

The 3-Element Algorithm techniques and results of a performance evaluation are presented. Based on the amount of forecast improvement, eight configurations were retained for analysis across the entire dataset containing six severe weather cases. A new stability proxy, the Elevated Total-Totals Index, was developed and integrated into the algorithm to improve severe weather forecasts over high-elevation regions where some traditional severe weather indices cannot be accurately computed. Additionally, the horizontal gradient of convective available potential energy was studied as a new indicator to the presence of dynamic forcing. It is hoped that improvements discussed in this paper will make the 3-Element Algorithm an effective tool in the early forecasting of severe weather, increasing lead-time to safeguard lives and resources.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my thesis advisor, Professor Wendell Nuss and second reader, Dave Brown. Professor Nuss provided a direction for the thesis as well as constant guidance. His tireless efforts to convert data from multiple formats to a single format for the experiments, expertise on the Linux operating system and the FORTRAN programming language were pivotal to the completion of the thesis. Mr. Brown provided insight from a new perspective, which helped answer important questions that were not originally included in the thesis. Additionally, I would also like to thank Bob Creasey. He ensured computer and network resources were made available, and these systems functioned properly during the course of this project.

I would also like to thank my wife Dina for her continuous support through the completion of this degree.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Severe convective weather poses a serious threat to the safety of people and resources. Research over the past 50 years has increased our understanding of the origins of severe weather, which has led to improved forecasts. Through this research, several indices were developed that attempt to aid forecasters in forecasting the threat for severe thunderstorms. Despite decades of research, there is still no “one size fits all” severe thunderstorm threat index or forecast product whose forecast skill is independent of geographic region. This is primarily due to the fact that there are many variables that contribute to the development of severe thunderstorms: instability, dynamic forcing, moisture, vertical wind shear, terrain, local effects, and tertiary circulations. Adding to these challenges is the fact that severe convection occurs on space and time scales that are often too small for most computer models to accurately resolve and forecast.

In an attempt to develop a universal severe weather forecast tool, the Air Force Weather Agency (AFWA) developed a 3-Element Severe Weather Forecast Algorithm (hereafter referred to as the 3E Algorithm) which is running on the 45 km MM5 model (Keller 2004). The 3E Algorithm generates a "yes/no" severe weather forecast based on atmospheric stability, cap strength, and the presence of a trigger using several commonly-applied severe weather indices. While not intended to pinpoint exactly when and where a severe thunderstorm will develop, the 3E Algorithm is intended to alert the meteorologist to areas where the potential for severe convection exists.

The main limitation of the 3E Algorithm is its inability to determine the type of severe weather expected, as the algorithm does not consider the specific atmospheric conditions that affect the evolution of severe thunderstorms. Studies have shown that the type of conditions produced by severe thunderstorms is highly dependent on the amount of vertical wind shear relative to convective instability and/or dynamic forcing (Rotunno 1984; Rasmussen and Blanchard 1998).

The objectives of this study were to 1) document the performance of the 3E Algorithm in its original configuration, thereby establishing a baseline upon which to improve; 2) develop a set of configurations that could potentially improve performance

above the original configuration, and 3) determine the optimal configuration for the 3E Algorithm for the 32 km NAM model through a two-tiered experimental approach.

II. BACKGROUND

The AFWA 3E Algorithm was developed in 2004 to identify favorable regions for the development of severe weather by considering the presence of three elements: instability, weak or no cap, and dynamic forcing. It is currently running as an experimental product on the AFWA 45 km MM5 model. The AFWA version of the 45 km MM5 model is 192 by 139 gridpoints horizontally and contains 41 vertical levels using the 3DVAR data assimilation and analysis scheme. It also includes the Grell cumulus parameterization scheme in an attempt to account for sub-gridscale updrafts. Additional information on the MM5 characteristics is available via the AFWA website at https://weather.afwa.af.mil/HOST_HOME/DNXM/ABOUTMM5/toc/physics/index.html.

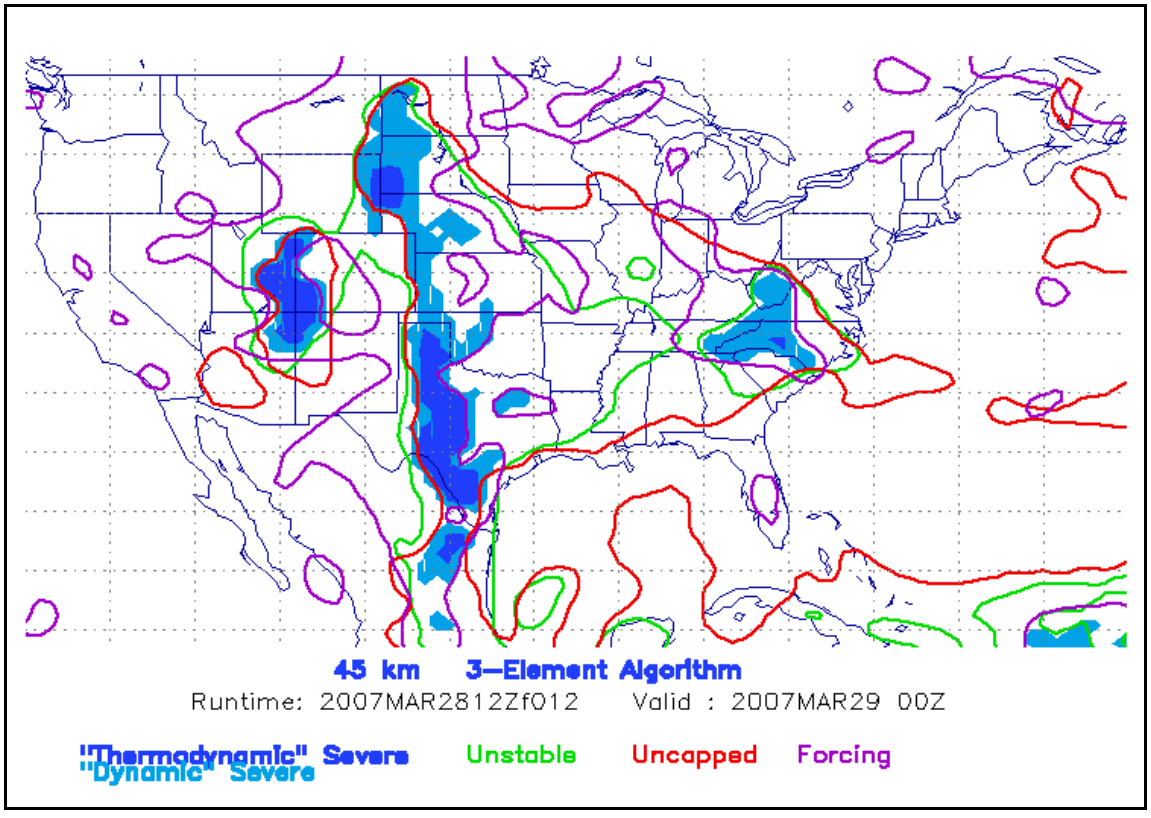

Figure 1. The 3-Element Severe Weather Forecast Algorithm on the 45 km MM5 model. Colored lines indicate areas where each of the three severe weather elements are present. The blue shaded areas are where all three elements are present and the algorithm is forecasting severe thunderstorms. [From Air Force Weather Agency]

The algorithm functions as a series of Boolean operators that checks for the presence of each of the three severe weather elements through a series of atmospheric indices that act as proxies for each element. Once the 3E Algorithm has determined the presence of all three elements at a given gridbox and time, it flags that gridbox as "yes" for severe weather potential. Otherwise, the algorithm flags the gridbox as "no" for severe weather if one or more elements are missing. The following section describes each of the elements, their associated proxies and thresholds. Equations for computing the proxies are listed in Appendix A.

A. ALGORTHM ELEMENTS AND THRESHOLDS

1. Convective Instability

A potentially unstable atmosphere is almost absolutely necessary for the initiation of convection, let alone severe thunderstorms. The algorithm checks for atmospheric stability by considering the convective available potential energy (CAPE), Total Totals Index (TT), and Lifted Index (LI). Empirical results (Miller 1972), established the minimum CAPE threshold at 3000 J kg^{-1}, the Total Totals threshold at 55, and the Lifted Index threshold at -5.0. The gridbox is considered unstable if one or more of the proxies meets its respective threshold.

2. Cap Strength

Continuing along the lines of instability, any inversion (cap) within the column will act to inhibit convection. However, a weak cap that can be penetrated by warm, saturated (or near-saturated) air parcels given sufficient surface heating or dynamic forcing can actually enhance the severity of thunderstorm convection. Analogous to removing the lid from a pot of water heated above the boiling point, some of the most violent severe weather events began with moist, unstable air being forced through a weak cap, resulting in explosive convective development through sudden condensation and latent heat release. The 3E Algorithm checks for convective inhibition values meeting the required threshold of 50 J kg^{-1} or less, and Lid Strength Index (LSI) of 2 K or lower. Both of these conditions must be met for favorable cap strength to be present.

3. Dynamic Forcing

Forcing by a triggering mechanism is necessary to initiate upward motion of air parcels in a conditionally unstable column. The trigger can come in many forms: sufficient surface heating, boundary layer convergence, surface fronts, divergence aloft, and outflow boundaries from nearby thunderstorms are examples of triggering mechanisms that often help initiate severe thunderstorm development. The 3E Algorithm checks for an upward vertical velocity at 700 mb, positive boundary layer convergence, and warm advection at 850 mb. Once the thresholds for these three proxies are met, the algorithm indicates the presence of dynamic forcing.

It is suggested that extreme dynamic forcing may initiate severe convection in lieu of convective instability (Harnack et. al. 1997), given a number of severe weather events that occurred due to very strong forcing in an otherwise relatively stable environment. Therefore, it has been proposed that the convective instability element may be weighted less, and possibly ignored entirely, when the dynamic forcing is strong enough to overcome any convective stability. One of the cases examined in this study occurred in a relatively stable environment: the February 12, 2007 tornado event in Louisiana and Mississippi.

B. CONDITIONS EXCLUDED FROM THE 3E ALGORITHM

The 3E Algorithm is capable of alerting the forecaster to regions favorable for severe weather based on cap strength, instability, and dynamic forcing. However, the type of conditions produced by a severe thunderstorm is dependent on vertical wind shear in relation to the buoyancy of the air mass (Harnack et. al. 1997). As a result, the Bulk Richardson Number (BRN) is useful in differentiating between single-sell, multi-cell, and supercell thunderstorms (Rotunno 1984). Still, there are some limitations to using the BRN in assessing the type of severe weather threat, particularly in relatively low CAPE environments where severe thunderstorms are initiated by extreme dynamic forcing. Complementing the BRN, Thompson et al. (2005) suggest that using the storm-relative helicity (SRH) may overcome some of the limitations in using the BRN for forecasting the potential for tornadic supercell development.

C. LIMITATIONS OF THE 3E ALGORITHM

A major limitation to the 3E Algorithm in its present form is it indicates regions where favorable conditions for severe weather exist, but cannot specify the type of severe weather expected. Knowledge of the type of severe weather that could potentially impact a given location is critical in the decision-making process in determining how and when to take action to protect lives and resources. Differences in the relationship between vertical shear and convective instability determine whether severe convection will come in the form of single-cells, multi-cell clusters and lines, mesoscale convective systems (MCS) and complexes (MCC), tornadic and non-tornadic supercells (Rotunno 1984, Rasmussen and Blanchard, 1998). Two key components affecting the evolution of severe thunderstorms are the magnitude of the vertical shear, and the strength and orientation of the low-level jet (LLJ) with respect to the mean flow in the middle and upper levels of the atmosphere. A strong LLJ contributes moisture and warm air, while acting to increase the vertical shear through both its speed and direction.

Additional limitations of the algorithm include: overforecasts during the diurnal minimum time of severe weather; excludes surface temperature, dew point, and terrain effects; overforecasts over the Rocky Mountains and Pacific Northwest; and storm reports often occur northeast of the region in which the algorithm forecasts severe weather (Keller 2004). A particular challenge was improving the 3E algorithm to more accurately forecast severe convection over the relatively data-sparse and mountainous regions of the western United States. Severe weather events often go undetected in the West due to the remoteness of this region, and the rugged terrain limiting the usability of ground-based radars. More significantly, most of the traditional severe weather indices in their pre-existing form cannot be applied to most places in the Intermountain West due to the region's high elevations.

D. SEVERE WEATHER CRITERIA

Multiple agencies throughout the United States use different criteria in verifying severe weather events; criteria established by the Air Force and National Oceanic and Atmospheric Administration (NOAA) were used in identifying severe weather events. According to Air Weather Service Technical Report 200 (Miller 1972) and Air Force

Manual (AFMAN) 15-129, a severe weather event is defined by the occurrence of one or more of the following: tornadoes, winds 50 knots or greater, and hail 0.75 inch or greater diameter. However, Army and Air Force installations with Category I aircraft (Appendix B) have lower hail and wind thresholds for severe thunderstorm criteria (typically 0.50 inch diameter and 45 knots, respectively). These exceptions from AFMAN 15-129 are established by the respective base commander.

Lightning at or near the installation is also of particular concern, but forecasting lightning was excluded from this study since it is not included as criteria for severe conditions. The Air Force has two standards for verifying weather events: an event occurs on station as reported by the observer on duty (objective verification), and an event that occurs within 5 nautical miles of the aerodrome either reported by a nearby observing site or detected by remote sensing (subjective verification). AFMAN 15-129 further specifies the desired forecast lead time for tornadoes is 30 minutes, damaging winds is 2 hours, and large hail is 2 hours. For the purposes of this study, the severe weather threshold was 50 knots for winds and 0.75 inch diameter for hail in accordance with criteria specified by AFMAN 15-129. Verification was defined as a severe weather occurrence within a specified gridbox, regardless of the gridbox dimensions.

In accordance with the objectives stated in the previous section, the first task was to run the algorithm in its original configuration across a limited dataset to obtain performance data to establish a baseline upon which to improve the algorithm. Then a variety of different configurations were set up and run across the same limited dataset and compared against the default configuration. During the Tier I experiments, alternative configurations were developed with consideration to terrain, geographic region, season, and dynamics. The configuration that demonstrated the best performance, based on relative operating characteristic (ROC) and Kuiper Skill Score (KSS), were retained for analysis across the entire dataset during Tier II experiments. The remaining configurations were eliminated from further study. When the configuration with the second-highest ROC and KSS within a group was within 5% of the best-performing configuration, both configurations were retained. In such situations, the two configurations fall within the arbitrarily-defined statistical tolerance, thus a statistical tie

existed. Finally, the configuration with the highest performance statistics was determined to be the optimal configuration for the 3E Algorithm on the 32 km NAM model.

III. DATA AND METHODOLOGY

The 3E Algorithm was run on the 32 km NAM model and its performance was documented for six cases of severe weather over various regions of the contiguous United States (CONUS). The 32 km NAM (also known as the NCEP Weather Research and Forecasting (WRF) model) has a horizontal domain of 349 by 277 gridpoints and 43 vertical levels.

Determining the optimal algorithm configuration was performed using a two-tiered approach. During the Tier I experiments, a total of 30 threshold settings across all six configuration groups were developed and experimented using model forecasts from three cases. To minimize biasing, the three cases and forecast time used for the Tier I experiments were chosen at random. Following the completion of the Tier I experiments, the algorithm settings with the highest improvement in performance along with the default configuration (A1) were retained for Tier II experiments. The lower-performing experimental configurations were eliminated from consideration in Tier II. During the Tier II experiments each configuration was run on all model runs and forecast times for each of the six cases. The performance of each Tier II configuration was analyzed according to model run, forecast time, case, and overall performance. The two-tiered approach was chosen as a time-saving measure largely based on the assumption that the configurations that perform poorly on the limited Tier I cases will also perform poorly in Tier II, and can therefore be excluded as potentially optimal configurations for the 3E Algorithm.

A. SETTING UP THE ALGORITHM

The Air Force Weather Agency wrote the original 3E Algorithm in a programming language that is compatible with its Gridded Analysis and Display System (GRADS) software. The first task was to re-write the algorithm using a programming language that generated output files recognizable by the Linux operating system and the General Meteorology Package (GEMPAK) suite used to process and display meteorological data. Fortran 77 was the program language selected for running the 3E Algorithm, since programs for extracting raw data and repacking the processed data files

for use by GEMPAK already existed in this computer language. Writing the 3E Algorithm in Fortran 77 also ensured it could be easily integrated into the major forecast models as a postprocessing operation because the forecast models are built largely from Fortran code. The datasets used for the study were stored in gridded binary files (GRIB) containing entire forecast sets for a given model run time. A separate program extracted the parameters and forecast times needed to run the 3E Algorithm, and placed these data into data files organized by parameter and forecast time. The algorithm was then run by loading the data files for the specified model run and forecast time, which generated four new data files containing the output for the stability element, cap strength element, forcing element, and the severe weather indicator. The algorithm also performed intermediate operations such as computing derived parameters used for assessing severe thunderstorm potential. The verification program was written so that it read in the algorithm output, then read in the Storm Prediction Center (SPC) Storm Reports file and the WSR-88D composite analysis. The program then generated a verification data file for display in GEMPAK and computed the algorithm performance as described in Section B below. Using the GEMPAK Analysis and Rendering Program (GARP), the data files were loaded onto a map of the United States to visually compare where the algorithm forecasted severe weather to the locations where severe thunderstorms actually occurred. A number of tests were performed to ensure all elements of the algorithm and verification program functioned properly before experiments began.

B. SPATIAL VERIFICATION METHODOLOGY

Storm report data from the National Oceanic and Atmospheric Administration Storm Prediction Center was used for verifying 3E algorithm forecasts and assessing its overall performance. The SPC Storm Reports are a compilation of observed severe weather events reported by trained spotters, law enforcement officials, and the general public. However, these reports also include observations from official National Weather Service observation sites, both manned and unmanned. While the storm reports were regarded as directly observed severe weather events, the impacted areas of these reports are very small, and large data gaps were the result. Therefore, WSR-88D data was also included to augment SPC Storm Reports by associating a minimum reflectivity threshold

to the occurrence of a severe thunderstorm. A base reflectivity of 40 dBZ was set as the minimum threshold for verifying the occurrence of severe weather using WSR-88D data. This threshold was selected based on the Probabilities of Detection (POD) of 70% for isolated severe thunderstorms, 64% for MCS/squall lines, and 82% for supercell thunderstorms (Johnson et. al. 1997). The main limitation to using WSR-88D data for verification was the lack of distinction between echoes from thunderstorms versus heavy stratiform precipitation or sleet, both of which may have had echo returns of 40 dBZ or higher. To mitigate these errors, a near-surface temperature of 18°C was also required for any gridbox containing radar echoes of 40 dBZ or greater, assuming that the potential of severe thunderstorms diminishes rapidly with temperatures below 18°C (Miller 1972).

	Observed Yes	Observed No
Forecast Yes	A	B
Forecast No	C	D

Figure 2. Contingency diagram used to verify 3E Algorithm forecasts. [From Fowle and Roebber 2003]

Using the contingency diagram in Figure 2 (Fowle and Roebber 2003), the forecast for each gridbox was compared against observed conditions. The desired state for each gridbox is for either a forecast value of "yes" and an observed value of "yes," or

a forecast value of "no" and an observed value of "no," (Bins A and D) which is defined as a verified forecast. A “false alarm” was the result of a forecast value of “yes” and an observed value of “no” (Bin B) while a missed forecast resulted from a forecast value of “no” and an observed value of “yes” (Bin C).

Once the forecast/observation comparison was completed for all gridboxes, the totals in Bins A, B, C, and D were used to compute the probability of detection, miss rate, false-alarm rate, bias, threat score, correct nonoccurrence, and Kuiper skill score. Equations used to compute these statistics are listed in Appendix C. The ROC area was also computed to compare the hit rate against the false alarm rate.

While the formulas in Appendix C are widely accepted for determining the overall performance of a given computer model or algorithm, they carry a major caveat requiring severe weather to have occurred somewhere within the model domain, or else some of the equations fail due to division by zero. During the verification process, algorithm outputs for gridpoints lying outside of the CONUS were disregarded, since there was no practical means of verifying the occurrence of severe weather over these areas. A simple check was included in the verification program to exclude non-CONUS gridpoints from the counts for calculating the algorithm’s performance. Furthermore, the study eliminated from consideration any gridpoints where the near-surface temperature was at 18º C or below, since severe thunderstorms require the presence of abundant potential energy in the form of warm surface temperatures and low-level moisture, indicated by a near-surface dew point of 18º C or greater (Miller 1972).

C. TEMPORAL TOLERANCES AND TIMING ERROR

The 3E Algorithm generates a “point-in-time” forecast product, where the algorithm output is a “snapshot” of the atmospheric state at the given forecast time. This methodology does not account for atmospheric conditions in between forecast times. Complicating matters is the fact that severe thunderstorms often occur on time scales small enough that the event falls in between forecast forecast times. To ensure continuity in the verification process, the forecast for a gridbox will verify if the forecast condition matches the observed condition in the time range of 90 minutes prior to the forecast valid time to 90 minutes following the forecast forecast time. The selected temporal tolerances

were consistent with the desired lead time criteria set forth by AFMAN 15-129. For example, if severe weather was forecast for a given gridbox at 1200Z and it occurred at 1345Z, the verification was recorded as a “false alarm” for that gridbox since it occurred 1 hour and 45 minutes later than its forecast time. Similarly, if severe weather was forecast at 1200Z and it occurred at 1020Z, a missed forecast for that gridbox was recorded because it occurred 1 hour and 40 minutes earlier than forecast. After the storm report and radar data were overlaid on the 3E Algorithm forecast map, the number of gridboxes containing verified forecasts, "false alarms," and missed forecasts were tallied to compute the POD, false alarm rate (FAR), algorithm bias, threat score (TS), Kuiper Skill Score (KSS) and relative operating characteristic (ROC).

D. CONTROL AND EXPERIMENTAL GROUPS

The 3E Algorithm was run on the 32 km NAM model in six different configuration groups, each based on hypothetical conditions that would favor the development of severe thunderstorms. Configuration group A contained the original set of severe weather indices used as proxies. The remaining five configuration groups contained additions, subtractions, or regional modifications to the original set of indices. Within each configuration group are a number of configurations, which are different threshold settings for each proxy used by that configuration group.

The control group was the algorithm configuration used by AFWA at the time this study was conducted, referenced hereafter as the default setting. Configurations modified from the default settings were the experimental groups. The following tables present a description of each algorithm configuration used during the study. For reference, “TT” refers to the Total-Totals Index; “ETT” is the Elevated Total-Totals Index; “LI” is the Lifted Index; “LSI” is the Lid Strength Index; “CIN” is the convective inhibition; “850TAD” is the thermal advection at 850 mb; “700TAD” is the thermal advection at 700 mb; “VV700” is the 700 mb vertical velocity (omega); “GCAPE” refers to the horizontal CAPE gradient; and “CNVG” is the surface moisture convergence.

Config	TT	LI	CAPE	LSI	CIN	850TAD	VV700	CNVG
A1*	55	-5.0	3000	2.0	-50	< 0	< 0	> 0
A2	50	-4.0	3000	2.0	-50	< 0	< 0	> 0
A3	52	-4.0	3000	2.0	-90	< 0	< 0	> 0
A4	50	-4.0	3000	2.0	-90	< 0	< 0	> 0
A5	50	-3.5	2800	2.0	-90	< 0	< 0	> 0
A6	50	-3.0	2800	2.0	-90	< 0	< 0	> 0
A7	50	-3.0	2600	2.0	-90	< 0	< 0	> 0

Table 1. Algorithm configuration group A: the original set of 3-element proxies established by AFWA. The configuration followed by an asterisk is the control group.

Config	TT/ETT	LI	CAPE	LSI	CIN	850/700 TAD	VV700	CNVG
B1	55/30	-5.0	3000	2.0	-50	< 0	< 0	> 0
B2	50/30	-5.0	3000	2.0	-50	< 0	< 0	> 0
B3	50/32	-4.0	3000	2.0	-50	< 0	< 0	> 0
B4	50/32	-4.0	2800	2.0	-90	< 0	< 0	> 0
B5	50/30	-3.5	2800	2.0	-90	< 0	< 0	> 0

Table 2. Algorithm configuration group B, using modified proxies with 700 mb thermal advection (700 TAD) and Elevated Total-Totals (ETT) over the Intermountain West.

Configuration groups B and F experimented with regionalizing the 3E Algorithm, with modified indices proposed for high-elevation areas such as the Intermountain West. This region is graphically outlined in Figure 3.

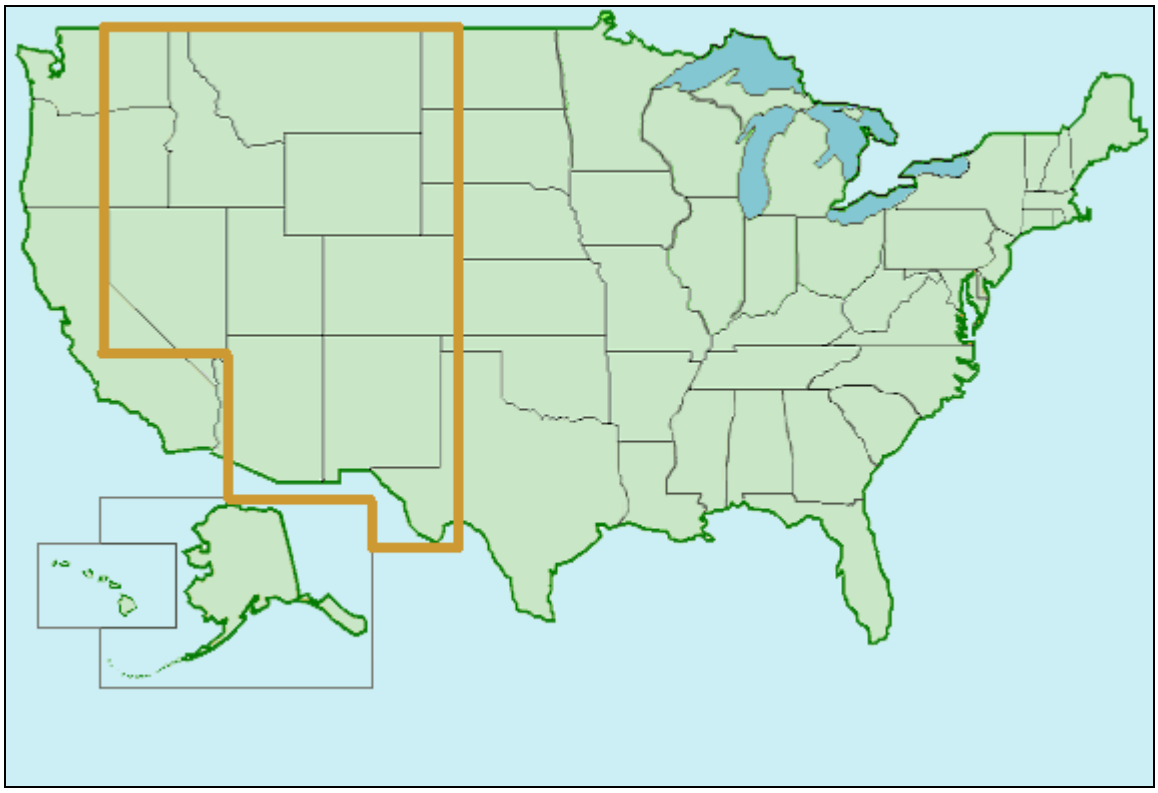

Figure 3. The outlined region is the Intermountain West used for configuration groups B and F.

Config	TT	LI	KI	CAPE	LSI	CIN	850 TAD	VV700	CNVG
C1	55	-5.0	30	3000	2.0	-50	< 0	< 0	> 0
C2	50	-4.0	30	3000	2.0	-50	< 0	< 0	> 0
C3	50	-3.5	30	2800	2.0	-90	< 0	< 0	> 0
C4	50	-3.5	28	2800	2.0	-90	< 0	< 0	> 0
C5	50	-3.2	27	2800	2.0	-90	< 0	< 0	> 0
C6	50	-3.0	27	2600	2.0	-90	< 0	< 0	> 0

Table 3. Algorithm configuration group C, which includes the K-Index (KI) as an additional stability proxy.

Config	TT	LI	CAPE	LSI	CIN	850 TAD	VV700	CNVG	GCAPE
D1	50	-5.0	3000	2.0	-50	< 0	< 0	> 0	1.0
D2	50	-4.0	3000	2.0	-50	< 0	< 0	> 0	1.0
D3	50	-3.5	2800	2.0	-90	< 0	< 0	> 0	1.0
D4	50	-3.0	2800	2.0	-90	< 0	< 0	> 0	1.0

Table 4. Algorithm configuration group D, which requires one or more of the four listed proxies for dynamic forcing.

Config	TT	LI	CAPE	LSI	CIN	850 TAD	VV700	CNVG	GCAPE
E1	55	-5.0	3000	2.0	-50	< 0	< 0	> 0	1.0
E2	50	-4.0	3000	2.0	-50	< 0	< 0	> 0	1.0
E3	50	-3.5	2800	2.0	-50	< 0	< 0	> 0	1.0
E4	50	-3.0	2800	2.0	-90	< 0	< 0	> 0	1.0

Table 5. Algorithm configuration group E, which requires a strong horizontal CAPE gradient and one or more of the three remaining proxies for dynamic forcing.

Unlike the previous configuration groups, which examine the effects of one proxy being added or removed at a time, configuration group F attempts to depict the cumulative effects of multiple proxies being added or removed, including the regionalization scheme used in configuration group B. It is anticipated this approach will constrain the forecast solution to more accurately represent the areas where severe thunderstorms are most likely to develop.

Config	TT/ETT	LI	KI	CAPE	LSI	CIN	850/700 TAD	VV700	CNVG	GCAPE
F1	55/30	-5.0	N/A	3000	2.0	-50	< 0	< 0	> 0	1.0
F2	50/30	-4.0	N/A	3000	2.0	-50	< 0	< 0	> 0	1.0
F3	50/30	-3.5	N/A	2800	2.0	-90	< 0	< 0	> 0	1.0
F4	50/30	-3.5	30	2800	2.0	-90	< 0	< 0	> 0	1.0
F5	50/30	-3.5	30	2800	2.0	-90	< 0	< 0	> 0	1.0

Table 6. Algorithm configuration group F, using ETTI and 700 mb thermal advection over the Intermountain West and includes K-Index and horizontal CAPE gradient elsewhere.

1. High-terrain Indices for Mountainous Areas

While there are numerous mountain ranges with peaks topping 14,000 feet in the West, most of these ranges are separated by tens to hundreds of kilometers by high-level and relatively flat plateaus and basins that range between 4,000 and 5,000 feet in elevation. With that said, much of this region lies above the 850 mb level that serves as the lower boundary for the Total-Totals and K-Index. This high elevation also makes it impracticable to use the 850 mb thermal advection to determine the presence of dynamic forcing. To address the issues surrounding the Total-Totals Index and 850 mb thermal advection over the Intermountain West, modifications to these indices were developed and integrated into the algorithm as Configuration B. The 700 mb thermal advection was used in lieu of the 850 mb thermal advection as a dynamic forcing proxy. The Elevated Total-Totals Index was developed following the same general formula as the Total-Totals Index, except 700 mb was used as the lower bounds instead of 850 mb, and 500 mb was retained for the upper bounds. Analysis on several severe weather events over the Intermountain West (including but not limited to the Salt Lake City tornado event of 11 Aug 1998) indicates that an Elevated Total-Totals Index of 30 or greater is favorable for severe thunderstorms. The original set of proxies was used elsewhere.

2. K-Index

Configuration Group C included the K-Index as an additional stability proxy. The K-Index considers the vertical temperature and moisture profile between 850mb and 500mb, while also considering the amount of moisture present at 700mb. This index is particularly useful in determining the likelihood for the formation of thunderstorms based on convective instability. An index value of 26 indicates a thunderstorm potential of 50%, while a value of 30 or greater corresponds with an 85% likelihood of thunderstorms. The K-Index alone however is insufficient in determining whether or where conditions are favorable for severe thunderstorms versus ordinary convection. Combining the K-Index with other indices can alert the meteorologist to areas where conditions are favorable for severe thunderstorms and the likelihood of such thunderstorms exists.

3. Horizontal CAPE Gradient

CAPE is a thermodynamic quantity that accounts for the amount of energy that could be potentially released by a saturated air parcel rising through the atmosphere. It takes into account the environmental lapse rate, moisture content of the parcel, and the temperature of the parcel before, during, and after it is lifted. As a result the CAPE is generally proportional to the temperature and moisture content of the parcel before it is lifted, and the environmental lapse rate. It is suggested that severe convection will occur in areas where a strong CAPE gradient is present, since a large change in CAPE over a relatively small distance can represent a strong thermal and/or moisture discontinuity often associated with forcing mechanisms such as fronts and drylines. To test this hypothesis the horizontal CAPE gradient was added as another dynamic forcing proxy in three algorithm configurations. Experimental results indicated a horizontal CAPE gradient of 1 J kg^{-1} km^{-1} or greater represented sufficient dynamic forcing for severe thunderstorms to occur. Configuration group D required one or more of the following: a CAPE gradient of 1 J kg^{-1} km^{-1} or greater, upward (negative) vertical velocity at 700 mb, warm advection at 850 mb, or surface moisture convergence. However, configuration group E required a horizontal CAPE gradient of 1 J kg^{-1} km^{-1} or greater, in addition to either upward vertical velocity at 700 mb, warm advection at 850 mb, or the presence of surface moisture convergence.

4. Multiple Indices

Configuration group F is an integration of the modified stability and forcing parameters (used in configuration group B) over the Intermountain West, and the horizontal CAPE gradient and K-Index elsewhere within the model domain under different conditions. Configurations F1 through F4 included the horizontal CAPE gradient in the same setup as in configuration group D. That is, dynamic forcing was determined to be present if either the CAPE gradient was 1 J $kg^{-1}km^{-1}$ or greater, upward vertical velocity was present at 700 mb, warm advection at 850 mb, or there was surface moisture convergence. Like configuration group E, configuration F5 required a CAPE gradient of 1 J kg^{-1} km^{-1} or greater in addition to either upward vertical velocity at 700 mb, warm advection at 850 mb, or surface moisture convergence. Finally, configurations

F1 through F3 excluded the K-Index as a stability proxy, while the K-Index was included in configurations F4 and F5.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. DATA ANALYSIS AND RESULTS

A. CASES STUDIED

Six severe weather events over various geographical regions during early 2007 were examined. Several factors were common among all of the cases, such as an amplified longwave pattern, the presence of a strong baroclinic zone separating warm, moist air from cooler, drier air, and cyclogenesis in response to one or more shortwave troughs moving through the base of the upstream longwave trough. The first case was unique in the fact that severe thunderstorms occurred on both sides of the warm front as the surface low approached southern portions of Louisiana, Mississippi, Alabama, and western Florida from the northern Gulf of Mexico. The following describes the synoptic conditions that contributed to each event; no charts are presented.

1. February 1-2, 2007

This tornado and damaging wind event occurred across central Florida between 01/2000Z and 02/1200Z, resulting in 19 deaths and five injuries. Surface and upper-level reanalysis from 01/0000Z to 02/1200Z depicted a split flow and a deep trough over the Rocky Mountains at 300 mb. The polar front jet (PFJ) originated in western Canada, and then went south along the West Coast before dissipating over northern Arizona just upstream of the long-wave trough axis. East of the trough axis the jet reformed over southern New Mexico, combining with the subtropical jet (STJ) to create a large "shared-energy" region from southern Texas to the Carolinas. The combined jets supported a baroclinic zone stretching from western Texas to Florida that was manifested at the surface in the form of a quasi-stationary front. The 01/0000Z analysis showed a 130-knot jet streak over western Texas that intensified to 160-knots over Mississippi by 02/0000Z. This jet streak interacted with the baroclinic zone in the Gulf of Mexico, initiating surface cyclogenesis near New Orleans, Louisiana by 01/1200Z. Over the ensuing 24 hours, the low paralleled the Gulf Coast before turning northeast near Pensacola, Florida. The southerly flow ahead of the low center advected a deep layer of warm, moist air from the Caribbean, contributing to the formation of a low-level jet (LLJ) at the 850 mb level over

central Florida. Meanwhile another surface low formed on the boundary south of Louisiana in response to another minor shortwave trough in the upper-levels situated along the Texas/Louisiana border. As the first low center deepened over southern Georgia, the LLJ at 850 mb had strengthened to 50-knots, increasing the deep-layer shear and convective instability over central Florida. By 02/1200Z the primary low pressure center was located about 100 NM southeast of Cape Hatteras, North Carolina, while the second low pressure center had moved northeast to near Tallahassee, Florida. Reanalysis of the synoptic charts for February 1 and the morning of February 2 suggest that much of the severe weather occurred in response to a surface trough trailing from the first surface low. The second low pressure system that later pulled the frontal boundary through Florida had little severe convection associated with it, since the first system's thunderstorms sufficiently stabilized the atmosphere over Florida to inhibit the formation of new convection following the initial line of supercells.

2. February 12-14, 2007

The system responsible for producing 31 tornadoes, resulting in one death and 48 injuries formed east of the Rocky Mountains over northeastern New Mexico in response to a major shortwave trough moving east out of Arizona. The trough was supported by divergence from the left-exit region of a 130-knot jet max at 300 mb over northern Mexico. The main baroclinic zone was situated from Montana, southeastward along the east side of the Rocky Mountains to southeastern Colorado, before turning east across Oklahoma and into Arkansas. A dryline formed over central Texas due to moisture advection ahead of the low and the intrusion of dry air from the Mexican Plateau. When the primary low formed on 12 February, cold advection along the leeside of the Rocky Mountains helped to amplify the longwave trough and shifted its axis to the Mississippi Valley. As the developing low and supporting upper-level features ejected eastward across the southern Plains, severe convection developed over southeast Texas as the strengthening LLJ advected subtropical moisture and increased the deep-layer shear ahead of the advancing dryline. As the system moved to the northeast over the ensuing 24 hours, a line of tornadic supercells developed along a pre-frontal trough as it swept eastward along the Gulf Coast. Finally, as the parent low moved into the Ohio Valley,

secondary cyclogenesis occurred over South Carolina around 14/0000Z, as the upper-level shortwave encountered a strong baroclinic zone associated with cold-air damming east of the Appalachian Mountains. This transfer of energy from the primary low in the Ohio Valley to the developing circulation on the coast resulted in re-intensification of the LLJ leading to additional severe thunderstorms over the Carolinas from 13/1800Z to 14/0300Z. Beyond this time period the new surface low over the Carolinas rapidly intensified as it moved up the East Coast, resulting in a severe snow and ice storm for the mid-Atlantic and New England.

3. February 23-25, 2007

The severe weather outbreak that began during the afternoon of February 23rd and continued until the early morning of February 25th was in response to a rapidly deepening low pressure system that had formed over southeastern Colorado by 23/0000Z, and moved northeast across the Great Plains to central Iowa by 25/1200Z. The low formed on the downstream side of a deep longwave trough situated over the Great Basin. A shortwave trough and supporting jet streak at 300 mb was analyzed off of the northern California coast at 23/0000Z. This piece of energy dropped southeastward, passing over extreme southern California and Arizona before turning eastward across New Mexico. This shortwave trough helped deepen the longwave trough, creating sufficient upper-level divergence over the baroclinic zone east of the Rocky Mountains to initiate surface cyclogenesis over southeastern Colorado. During cyclogenesis, a dryline formed from western Texas northward to southwestern Kansas as the LLJ advected maritime tropical air northward from the Gulf of Mexico. As the system moved to the northeast, severe thunderstorms formed ahead of the dryline over western Kansas, Oklahoma, and Texas during the afternoon and evening of February 23, and propagated to the east with the advancing cold front into Louisiana, Arkansas, and Missouri on February 24. Tornadic supercells propagated eastward with the cold front into Mississippi during the early morning hours of February 25. After 25/0600Z the event came to an end as severe convection began to dissipate as the surface low became vertically stacked and began filling (weakening) over Iowa. Over the course of three days the system spawned 29 tornadoes, resulting in 27 injuries and no deaths.

4. February 28-March 1, 2007

The synoptic setup that led to the tornado outbreak on February 28 and March 1 was similar to the previous case. The upper-level longwave trough was situated over the Rocky Mountains, with a major shortwave trough swinging through southern California, then following the U.S.-Mexico border to near El Paso Texas. Meanwhile, a strong baroclinic zone was situated along the leeward side of the Rocky Mountains. When the upper-level shortwave trough ejected northeast out of the southern Rockies, cyclogenesis occurred over southeastern Colorado due to upper-level divergence as the left-exit region of the 175 KT jet max moved over the baroclinic zone. As cyclogenesis occurred, a dryline formed from western Texas to central Oklahoma. Strong moisture advection occurred with the LLJ ahead of the dryline, while southwesterly flow advected dry air behind it. As the system moved northeast, severe thunderstorms and tornadoes formed from eastern Kansas and Oklahoma to Illinois on February 28. Additional severe convection developed along the triple-point of the system from Mississippi to the Carolinas, including the EF-5 tornado that destroyed a high school in Enterprise, Alabama, claiming nine lives. During the two-day event, there were 82 tornadoes resulting in 20 deaths and 28 injuries.

5. March 23-24, 2007

A deep trough was entrenched over the southern Rocky Mountains at 300 mb, which stacks nearly vertically to a closed low over northern New Mexico at 500 mb, 700 mb, and 850 mb. At the surface a stalled frontal boundary is situated from central Missouri westward through southern Kansas and into northern New Mexico. Upper-level analysis at 22/1200Z showed a 100 KT jet max moving through the base of the 300 mb trough over extreme southern Arizona, which helped to initiate surface cyclogenesis along the frontal boundary over northwestern New Mexico. Meanwhile the atmosphere over western Texas and eastern New Mexico had already begun to respond to the approach of the upper-level shortwave trough; a 30 KT LLJ had formed and was advecting moisture from the Gulf of Mexico northward from the Rio Grande Valley to northwest Texas at the 850 mb level. The surface low continued to develop as it moved east-northeast from northern New Mexico, roughly following the stalled frontal boundary

across the Texas panhandle and into southern Kansas. As the system moved east, the LLJ ahead of it strengthened to 45 KTS increasing the low-level wind shear, while the deep-layer shear also increased over the southern Plains with the 100 KT jet max approaching at 300 mb. Additional low-level moisture was advected into western Texas, eastern New Mexico, southern Kansas, and eastern Colorado. Unlike more classic severe weather cases across the Southern Plains, there was ample moisture at 700 mb for the duration of the event, and the expected dry-air intrusion at this level was largely absent. The surface low and supporting upper-level shortwave trough moved northeast across Kansas and into an area of less moisture and greater stability over Missouri, and the severe thunderstorm activity across the Southern Plains had ended by 24/1200Z. This event resulted in 133 severe weather reports including 20 tornadoes from northeastern Colorado to eastern New Mexico and western Texas, resulting in two injuries near the town of Clovis, New Mexico.

6. March 29-30, 2007

The longwave pattern was largely the same for the severe thunderstorm outbreak of 29-30 March 2007 as it was for the previous severe thunderstorm event of 23-24 March 2007. The deep upper-level trough remained in place over the southern Rocky Mountains. The main differences however, is the baroclinic zone associated with the stalled frontal boundary was much further north, running from the Iowa/Missouri border northwestward to the Black Hills of South Dakota. The 120 KT jet max and supported shortwave trough moving through the longwave trough were much stronger than with the previous event four days earlier, and cyclogenesis began to occur over central Colorado by 28/1200Z. This contributed to the formation of a dryline from the Texas/New Mexico border northward through extreme eastern Colorado and southeastern Wyoming. A stronger southwesterly flow advected warm, dry air from the Mexican Plateau at 700 mb, while a 40 KT south-southeasterly oriented LLJ formed ahead of the dryline over extreme western Texas, Oklahoma, Kansas and Nebraska. These elements in combination with the approaching 120 KT jet max ejecting northeastward out of the Southern Rockies created a highly favorable setup for the formation of severe thunderstorms, and more specifically tornadic supercells with a highly unstable

atmosphere and strong deep-layer shear. As the surface low ejected out of the central Rockies and into the Plains the system spawned 84 tornadoes from western Nebraska to northwestern Texas, resulting in six deaths and two injuries.

B. INITIAL SETUP, DEBUGGING, AND QUALITY CONTROL

A number of quality control analyses were performed during the coding and setup phase of the experiment. This involved running the 3E Algorithm and associated verification program on a number of test cases and writing the output for each proxy to a test data file. Locations of severe weather reports were recorded and the test data was analyzed for gridpoints in and around the location of the severe weather event. Finally, the test data was compared against actual data from the nearest Rawinsonde sites to ensure consistency between values computed by the algorithm and actual indices derived from the upper-air soundings.

This quality control resulted in changes to the way the moist adiabatic lapse rate was computed in the 3E Algorithm. Before testing, the moist adiabatic lapse rate was assumed to be 6° C km^{-1}, and was coded in the algorithm as such. Upon comparison with Rawinsonde data, it was determined that the moist adiabatic lapse rate does not remain constant with increasing height, rather it increases nonlinearly. Through the analysis of data from multiple Rawinsonde sites, it was estimated that the moist adiabatic lapse rate is around 4.5° C km^{-1} from the surface to 850 mb. It increases to roughly 5.0° C km^{-1} from 850 mb to 700 mb. The lapse rate further increases to near 5.7° C km^{-1} from 700 mb to 500 mb, and around 500 mb the moist lapse rate was estimated to be near 7.0° C km^{-1}. Above 500 mb the moist adiabatic lapse rate increases to the point where it begins to approach the dry adiabatic lapse rate since the colder, less dense atmosphere at these heights cannot retain the amount of water vapor that can be held in the low-levels. Because of this nonlinear variation, using the assumed moist lapse rate of 6° C km^{-1} led to significant errors in computing CAPE, CIN, and Lifted Index, which could have resulted in significant degradation of the algorithm's performance. Using the approximated lapse rate values for the respective pressure levels greatly mitigated these errors, resulting in more accurate estimations of CAPE, CIN, and Lifted Index.

C. TIER I EXPERIMENT RESULTS

Each configuration was tested on the 01 February 2007, 0000Z run, 18-hour forecast; the 24 February 2007, 0000Z run, 24-hour forecast; and the 23 March 2007, 1200Z run, 12-hour forecast. Tables in the following sections provide detailed performance information for each configuration and forecast time.

1. Configuration Group A

Changes in the index thresholds yielded identical performance statistics for configurations A2, A3, and A4 for the NAM 00Z, 01 Feb 2007, 18-hour forecast. These configurations were each run twice to confirm there were no errors due to user input that led to the identical results. The greatest operational (ROC) improvement occurred with configuration A6 at 0.5800, a ROC increase of 0.4680 from the default configuration A1. However, the performance of configuration A7 was very close (0.5775) to that of A6, with the KSS of A7 being the same as A6 at 0.509. From the standpoint of accuracy, which includes the correctly-forecasted non-events in addition to the correctly-forecasted hits, configurations A6 and A7 ended up in a tie at 87.6%, both surpassing the remaining five configurations. There were large variations in the performance statistics because this was a relatively isolated severe thunderstorm event with a fairly small number of storm reports and radar echoes meeting severe weather thresholds. Results for configuration group A are graphically presented in Figure 4.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
A1	4.2%	81.8%	0.229	0.350	-0.100	0.010	0.1120	83.4%
A2	27.1%	58.1%	0.646	0.197	0.102	0.207	0.3450	84.0%
A3	27.1%	58.1%	0.646	0.197	0.102	0.207	0.3450	84.0%
A4	27.1%	58.1%	0.646	0.197	0.102	0.207	0.3450	84.0%
A5	41.7%	50.0%	0.833	0.294	0.214	0.346	0.4585	85.5%
A6	58.3%	42.3%	1.208	0.406	0.339	0.509	0.5800	87.6%
A7	58.3%	42.8%	1.208	0.406	0.339	0.509	0.5775	87.6%

Table 7. Tier I experiment results using the 18-hour forecast from the 01 February 2007 0000Z NAM run for configuration group A.

Tier I performance data for configuration group A on the NAM 00Z, 24 Feb 07, 24-hour forecast is listed in Table 8. Like in the previous case, configurations A6 and A7 had both the highest ROC (0.6395 and 0.6380, respectively) and highest KSS (0.544 and 0.546, respectively) among the settings in group A. However, configuration A4 had the greatest accuracy (95.1%), followed by configurations A2 and A6 (both at 94.9%). Configuration A5 performed nearly as well as configurations A6 and A7, with a ROC of 0.6315, a KSS of 0.527, and 94.8% accuracy rate. Configuration A6 reflects a ROC improvement of 0.1850 over default configuration A1. Unlike the previous case, none of the configurations was run more than once, because each configuration yielded a unique forecast solution. This was largely due to the fact that this severe thunderstorm event was much larger in both spatial coverage and the number of severe thunderstorm reports and radar echoes meeting severe weather criteria (Figure 5). Therefore, small differences in the number of correctly-forecasted gridboxes had a lesser impact on the performance of each forecast solution.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
A1	21.6%	30.7%	0.311	0.197	0.047	0.196	0.4545	91.4%
A2	51.0%	28.1%	0.701	0.425	0.329	0.469	0.6145	94.9%
A3	41.1%	29.8%	0.576	0.350	0.237	0.374	0.5565	93.5%
A4	52.3%	27.6%	0.722	0.436	0.342	0.481	0.6235	95.1%
A5	58.5%	32.2%	0.863	0.457	0.373	0.527	0.6315	94.8%
A6	60.6%	32.7%	0.900	0.468	0.386	0.544	0.6395	94.9%
A7	61.0%	33.4%	0.917	0.467	0.386	0.546	0.6380	94.7%

Table 8. Tier I experiment results using the 24-hour forecast from the 24 February 2007 0000Z NAM run for configuration group A.

All configurations within group A performed statistically poor for the NAM 12Z, 23 Mar 2007, 12-hour forecast, with particular emphasis on the FAR ranging from 81% to 87% across all cases. Graphical output of each forecast solution for this event (Figure 6) depicted the forecasted area for severe thunderstorms stretching from southern Iowa westward to central Kansas, then southward through central Oklahoma and Texas. According to storm reports and radar imagery, severe thunderstorms occurred from

eastern Colorado and southwestern Kansas, southward along the Texas/New Mexico border region. Aside from that anomaly, configurations A6 and A7 once again had the highest ROC values, both at 0.4060. However, configuration A3 had the highest KSS at 0.352. Configurations A2, A4 and A5 also had higher KSS, all at 0.334, in contrast to 0.304 and 0.293 for configurations A6 and A7, respectively.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
A1	13.2%	81.7%	0.727	0.083	0.034	0.080	0.1575	88.2%
A2	64.8%	84.5%	4.184	0.143	0.127	0.334	0.4015	68.2%
A3	57.6%	81.4%	3.089	0.164	0.144	0.352	0.3810	76.0%
A4	64.9%	84.5%	4.184	0.143	0.127	0.334	0.4020	68.2%
A5	65.9%	85.3%	4.489	0.136	0.121	0.334	0.4030	65.9%
A6	67.1%	85.9%	4.780	0.131	0.117	0.304	0.4060	63.7%
A7	67.5%	86.3%	4.957	0.128	0.114	0.293	0.4060	62.3%

Table 9. Tier I experiment results using the 12-hour forecast from the 23 March 2007 1200Z NAM run for configuration group A.

The overall Tier I performance (Table 10) was derived by averaging the statistics for each configuration over all three cases. Consistent with the three cases analyzed, configuration A6 exhibited the highest ROC at 0.5419 and greatest improvement in KSS (0.357) over the default configuration A1. Therefore, configuration A6 was retained for Tier II analysis across the entire dataset for all six severe thunderstorm cases. However, configuration A7 was also retained for Tier II analysis. The ROC area for configuration A7 was 0.5405 and the KSS was 0.449, which results in a difference between configuration A6 of 0.0014 and 0.003, respectively. As a percentage, the ROC area difference between configurations A6 and A7 is 2.6%, while the KSS difference is 0.7%. Assuming an arbitrarily-defined ±5% margin of uncertainty, a statistical tie between configurations A6 and A7 resulted. Configurations A6 and A7 were also the only configurations to have a ROC area greater than the "no-skill" area of 0.5, thus reflecting an improvement in forecast skill. Configuration A5, with an overall ROC area of 0.4977, essentially forecasted with no skill, since its ROC area was 0.5% from the "no-skill" ROC area of 0.5. The remaining configurations had ROC areas that fell well below the

"no-skill" ROC area of 0.5, and eliminated from further analysis (except for configuration A1 which was retained as the control group).

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
A1	13.0%	64.7%	0.422	0.105	-0.006	0.095	0.2413	87.6%
A2	47.6%	56.9%	1.738	0.255	0.186	0.337	0.4537	82.4%
A3	41.9%	56.4%	1.437	0.237	0.161	0.311	0.4275	84.5%
A4	48.1%	56.7%	1.851	0.259	0.190	0.341	0.4568	82.4%
A5	55.4%	55.8%	2.062	0.296	0.236	0.397	0.4977	82.1%
A6	62.0%	53.8%	2.296	0.335	0.281	0.452	0.5419	82.1%
A7	62.3%	54.1%	2.307	0.334	0.280	0.449	0.5405	81.5%

Table 10. Overall Tier I performance for configuration group A.

Tier I output for each configuration in group A is depicted in the following figures. Verification data (red outline) is superimposed over algorithm forecast (green shaded, purple outlined regions) output for visual comparison of each configuration in the following figures. Charts are organized by case for side-by-side comparison of each configuration. While the algorithm output may show areas where it forecasts the potential for severe thunderstorms outside of the CONUS, these areas were not included when computing the performance statistics since no means of verifying the occurrence of severe thunderstorms over ocean regions or outside of the CONUS was readily available. The severe thunderstorm thresholds for the Total-Totals Index, CAPE, and CIN were lowered, which eased the constraint on the forecast solution. As a result, configurations A6 and A7 forecasted the potential for severe weather over a larger area that encompassed most of the region where severe thunderstorms occurred, which gave the latter two configurations a much stronger chance of verifying.

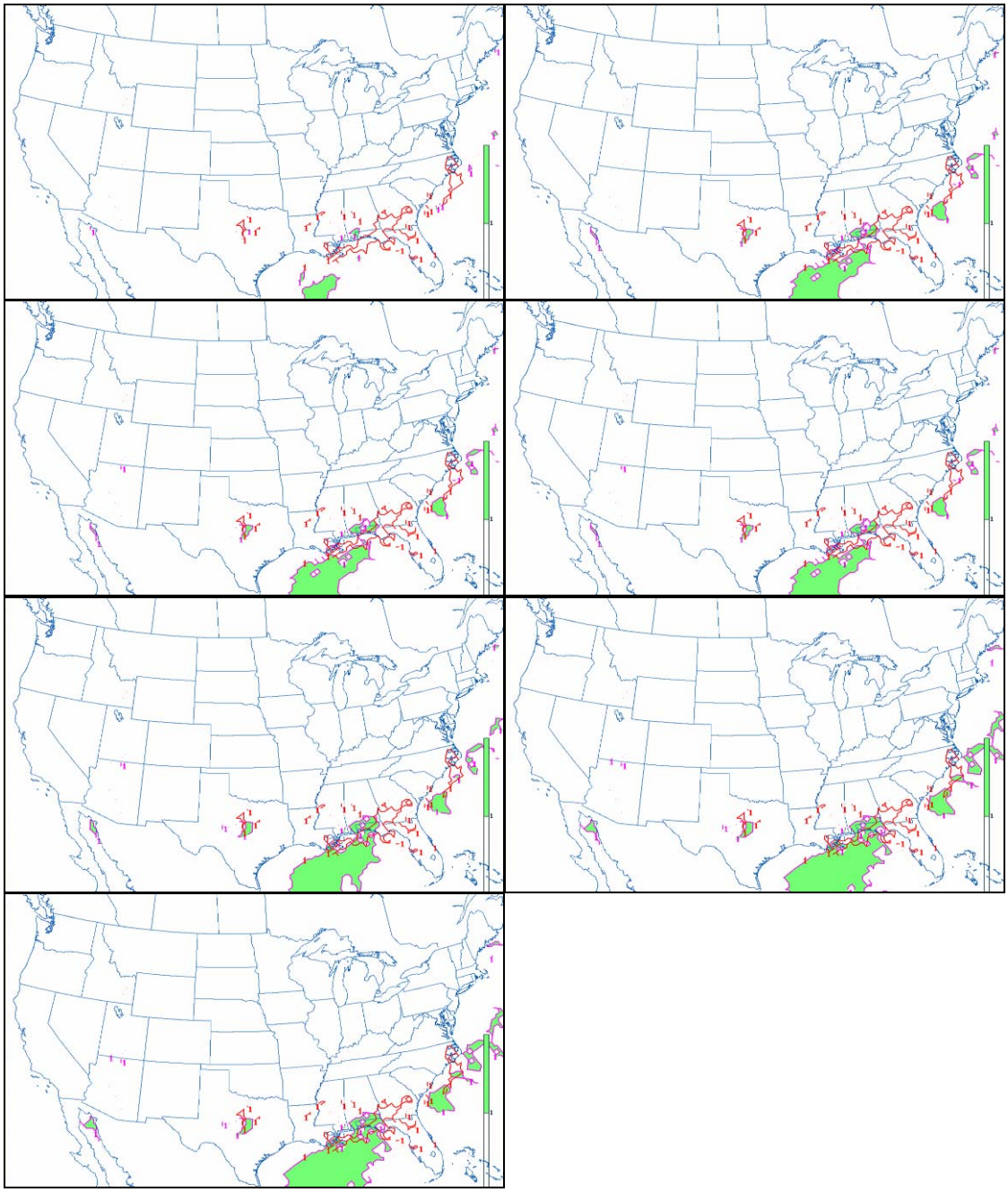

Figure 4. Configuration group A output for NAM 01 Feb 2007 0000Z run, 18-hour forecast (green shade, purple outline) and observed severe weather at 1800Z, 01 Feb 2007 (red outline). Configurations are ordered as follows: A1 (upper-left), A2 (upper-right), A3 (second row left), A4 (second row right), A5 (third row left), A6 (third row right), and A7 (lower left).

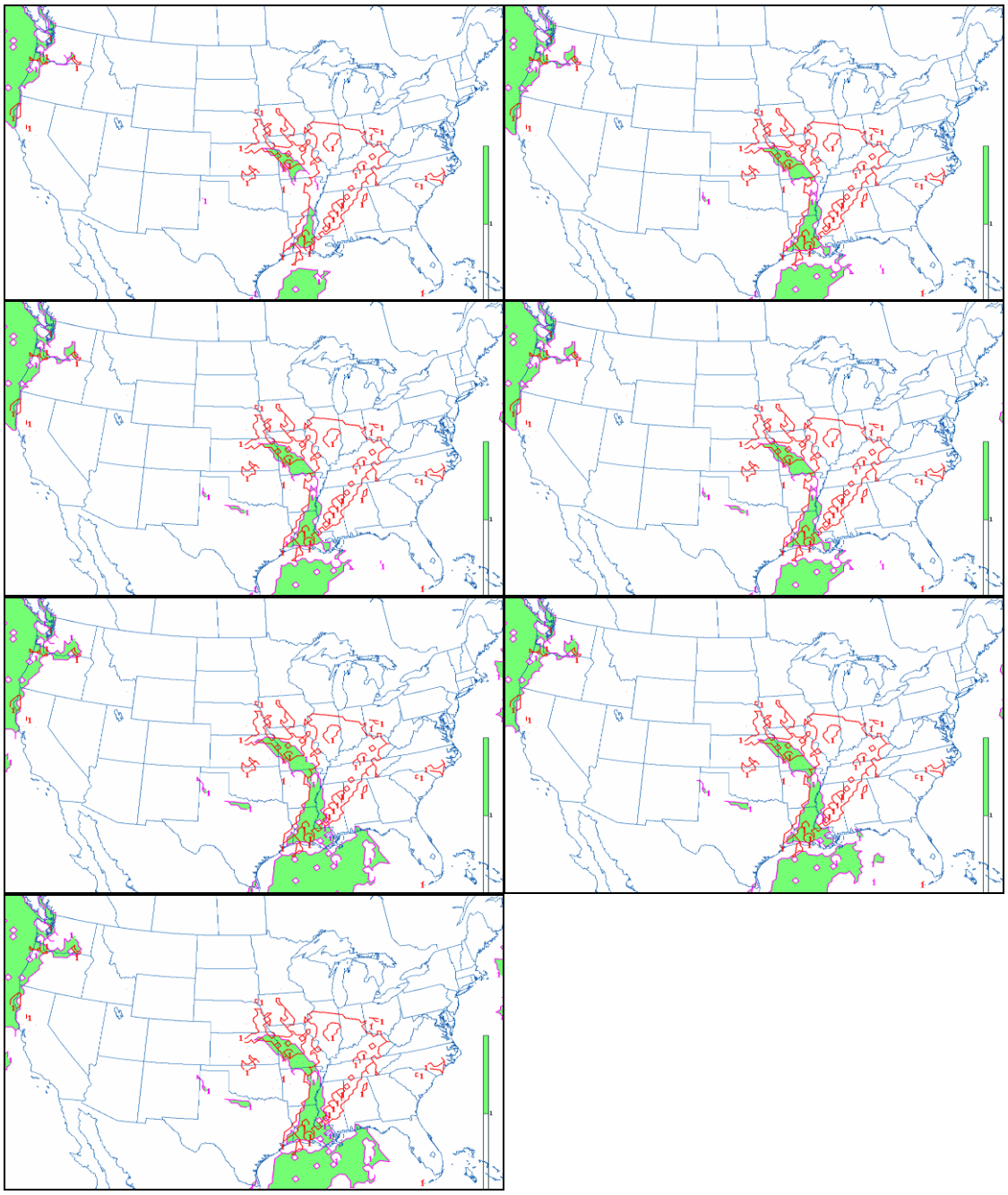

Figure 5. Configuration group A output for NAM 24 Feb 2007 0000Z run, 24-hour forecast (green shade, purple outline) and observed severe weather at 0000Z, 25 Feb 2007 (red outline). Configurations are ordered as follows: A1 (upper-left), A2 (upper-right), A3 (second row left), A4 (second row right), A5 (third row left), A6 (third row right), and A7 (lower left).

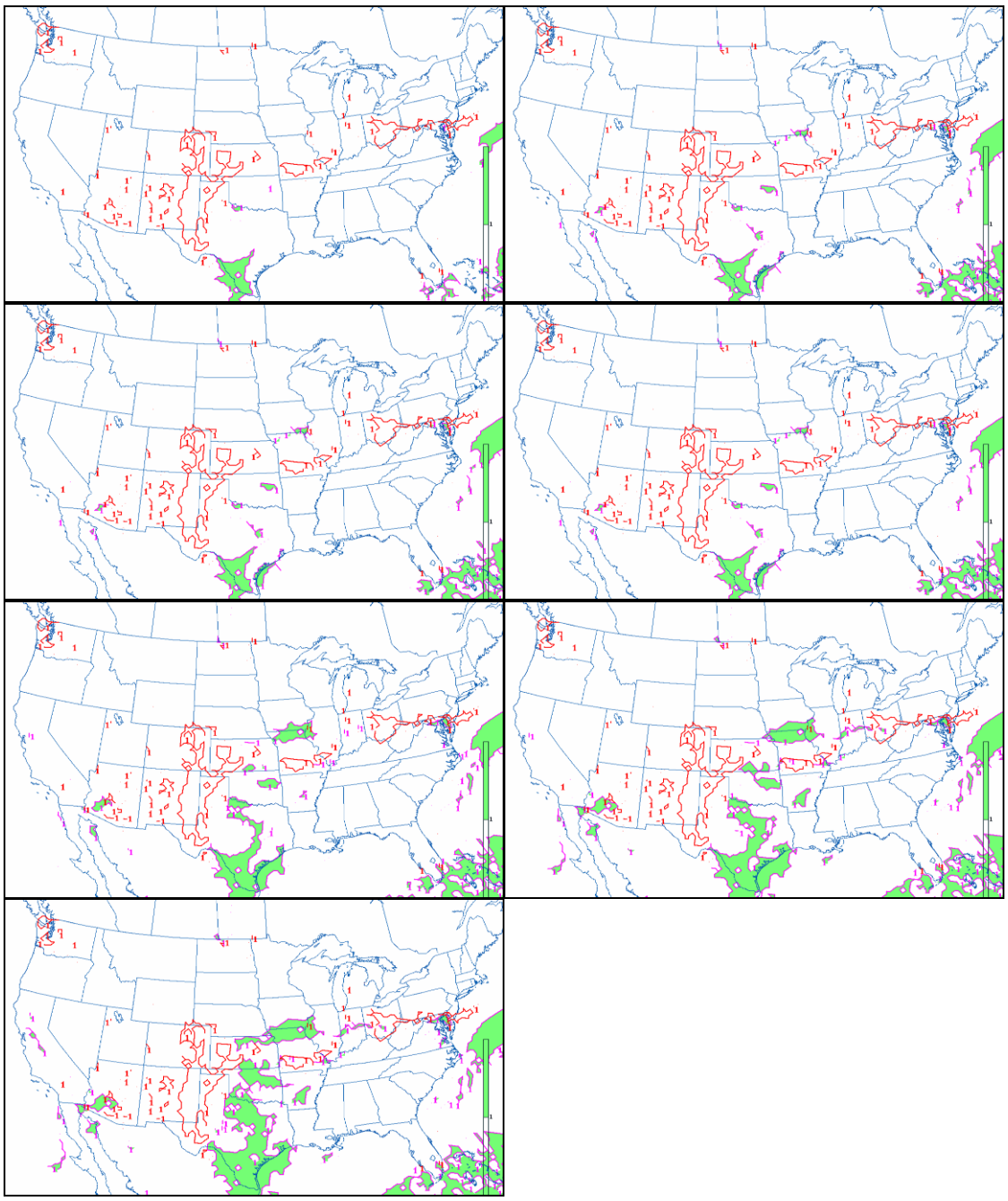

Figure 6. Configuration group A output for NAM 23 Mar 2007 1200Z run, 12-hour forecast (green shade, purple outline) and observed severe weather at 0000Z, 24 Mar 2007 (red outline). Configurations are ordered as follows: A1 (upper-left), A2 (upper right), A3 (second row left), A4 (second row right), A5 (third row left), A6 (third row right), and A7 (lower left).

2. Configuration Group B

Results for configuration group B were largely the same as for configuration group A for the NAM 00Z, 01 Feb 2007, 18-hour forecast as this event occurred mainly along the Gulf Coast. Nonetheless, the 3E Algorithm did not forecast severe thunderstorms over the Intermountain West using the Elevated Total-Totals Index and 700 mb thermal advection for any of the five configurations within this group. Therefore it was construed that the modified algorithm's forecasts resulted in a correctly-forecasted nonoccurrence of severe thunderstorms over the Intermountain West. Because the remainder of the country used the same set of proxies used in configuration group A, the results outside of the Intermountain West largely reflected those of the previous configuration group. For this particular case, configuration B5 had the highest ROC and KSS, at 0.4585 and 0.346, respectively. While configurations B2, B3, and B4, had different threshold settings, their performance statistics were identical due the isolated nature of this event, and the relatively few number of gridboxes in which severe thunderstorms were either forecasted or verified. Configuration B1, with the proxies set to the original AFWA thresholds outside of the Intermountain West, severely underforecasted the threat for severe thunderstorms in and around the area impacted. Numerical results are listed in Table 11, and graphically presented in Figure 7.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
B1	4.2%	81.8%	0.229	0.351	-0.991	0.010	0.1120	83.4%
B2	27.1%	58.1%	0.646	0.197	0.102	0.207	0.3450	84.0%
B3	27.1%	58.1%	0.646	0.197	0.102	0.207	0.3450	84.0%
B4	27.1%	58.1%	0.646	0.197	0.102	0.207	0.3450	84.0%
B5	41.7%	50.0%	0.833	0.294	0.214	0.346	0.4585	85.5%

Table 11. Tier I experiment results using the 18-hour forecast from the 01 Feb 2007 0000Z NAM run for configuration group B.

Performance data for configuration group B for the NAM 00Z, 24 Feb 07, 24-hour forecast is listed in Table 12, and graphically displayed in Figure 8. As in the previous case, configuration B5 performed optimally, compared to the remaining

configurations with a ROC of 0.6130 and a KSS of 0.517. However, while configurations B3 and B4 both had lower KSS values (0.469 and 0.482, respectively), both had slightly higher ROC areas than configuration B5 (0.6150 and 0.6180, respectively, versus 0.6130 for B5). With the exception of configuration B1, the ROC areas for all of the configurations exceeded the "no-skill" ROC area of 0.5.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
B1	21.6%	39.5%	0.357	0.189	0.045	0.186	0.4105	84.0%
B2	51.0%	32.4%	0.755	0.410	0.315	0.459	0.5930	87.3%
B3	51.1%	28.1%	0.710	0.426	0.329	0.469	0.6150	88.1%
B4	52.7%	29.1%	0.743	0.433	0.340	0.482	0.6180	88.1%
B5	58.5%	35.9%	0.913	0.441	0.357	0.517	0.6130	87.2%

Table 12. Tier I experiment results using the 24-hour forecast from the 24 Feb 2007 0000Z NAM run for configuration group B.

The algorithm forecasted the potential for severe thunderstorms over central Texas and Oklahoma for 00Z, 24 Mar 2007; these thunderstorms did not occur in this area, but several hundred kilometers to the west along the Texas/New Mexico border. Configuration B2 was the best-performing setting with a ROC area of 0.3495 and a KSS of 0.267. All configurations depicted the threat for severe thunderstorms over a large portion of the southern Rockies. The limited amount of available radar data indicated the occurrence of isolated potentially severe thunderstorms in the threat region. Data is presented numerically in Table 13, and graphically in Figure 9.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
B1	40.2%	83.2%	2.396	0.134	0.110	0.224	0.2850	78.8%
B2	55.2%	85.3%	3.739	0.132	0.114	0.267	0.3495	70.2%
B3	48.7%	85.0%	3.251	0.129	0.110	0.241	0.3185	73.2%
B4	34.3%	87.9%	2.838	0.098	0.076	0.121	0.2320	74.2%
B5	42.4%	88.1%	3.752	0.102	0.084	0.144	0.2715	69.5%

Table 13. Tier I experiment results using the 12-hour forecast from the 23 Mar 2007 1200Z NAM run for configuration group B.

Overall, Tier I experiments resulted in configuration B5 having the highest ROC and KSS (0.4475 and 0.4265, respectively), and was therefore retained for further analysis during the Tier II experiments. While configuration B5 was the top-performing setting in configuration group B, it performed poorer than configurations A6 and A7 (ROC area of 0.2715 versus 0.5419 and 0.5405, respectively). Outside of the Intermountain West, configuration A6 andA7 both performed somewhat better than configuration B5, because the Lifted Index threshold for both A6 and A7 was set at -3.0, while the same threshold was set at -3.5 for configuration B5. Because configuration A5 had identical threshold settings as B5 outside of the Intermountain West, the graphical results for B5 closely resembled those for A5 for the first two cases. That added constraint contributed to the aerial coverage for severe thunderstorms being underforecast in comparison to configurations A6 and A7, thereby resulting in a lower chance for verification. Poor performance in the third case lowered the overall performance statistics for configuration group B, largely due to the fact that the Elevated Total-Totals Index and 700 mb thermal advection resulted in the algorithm overforecasting the threat of severe thunderstorms in the West. However, this did not account for severe thunderstorms that could have potentially went undetected due to the sparse population and large gaps in radar coverage over the Intermountain West.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
B1	22.0%	68.2%	0.994	0.225	-0.279	0.140	0.2690	82.1%
B2	44.4%	58.6%	1.713	0.246	0.177	0.311	0.4290	80.5%
B3	42.3%	57.0%	1.536	0.251	0.180	0.306	0.4265	81.8%
B4	38.0%	58.4%	1.409	0.243	0.204	0.270	0.3980	82.1%
B5	47.5%	58.0%	1.832	0.279	0.218	0.336	0.4475	80.7%

Table 14. Overall Tier I performance for configuration group B.

While configuration B5 was retained for Tier II analysis, configurations B2 and B3 also performed relatively well. Neither of these configurations were retained for further study because while the ROC areas differences between B5 and both B2 and B3 fell within the ±5% tolerance for a statistical tie with configuration B5, the KSS differences of 0.025 for B2 and 0.030 for B3 both fell outside the allowed margin of error

(7.4% and 8.9%, respectively) to statistically tie configuration B5. Configurations B1 and B4 were the worst performing in group B, although they still performed better than the default configuration (A1). Forecast and verification data for configuration group B is displayed in the following figures. Consistent with the results in configuration group A, the configurations in group B with the lowest threshold settings had the least amount of constraint on the forecast solution, and therefore had the greatest spatial coverage for severe thunderstorms in the forecast. There was little difference in aerial coverage between configurations B5, A6, and A7 for the first two cases. The main differences occurred during the third case where severe thunderstorms were concentrated along the eastern edge of the Rocky Mountains.

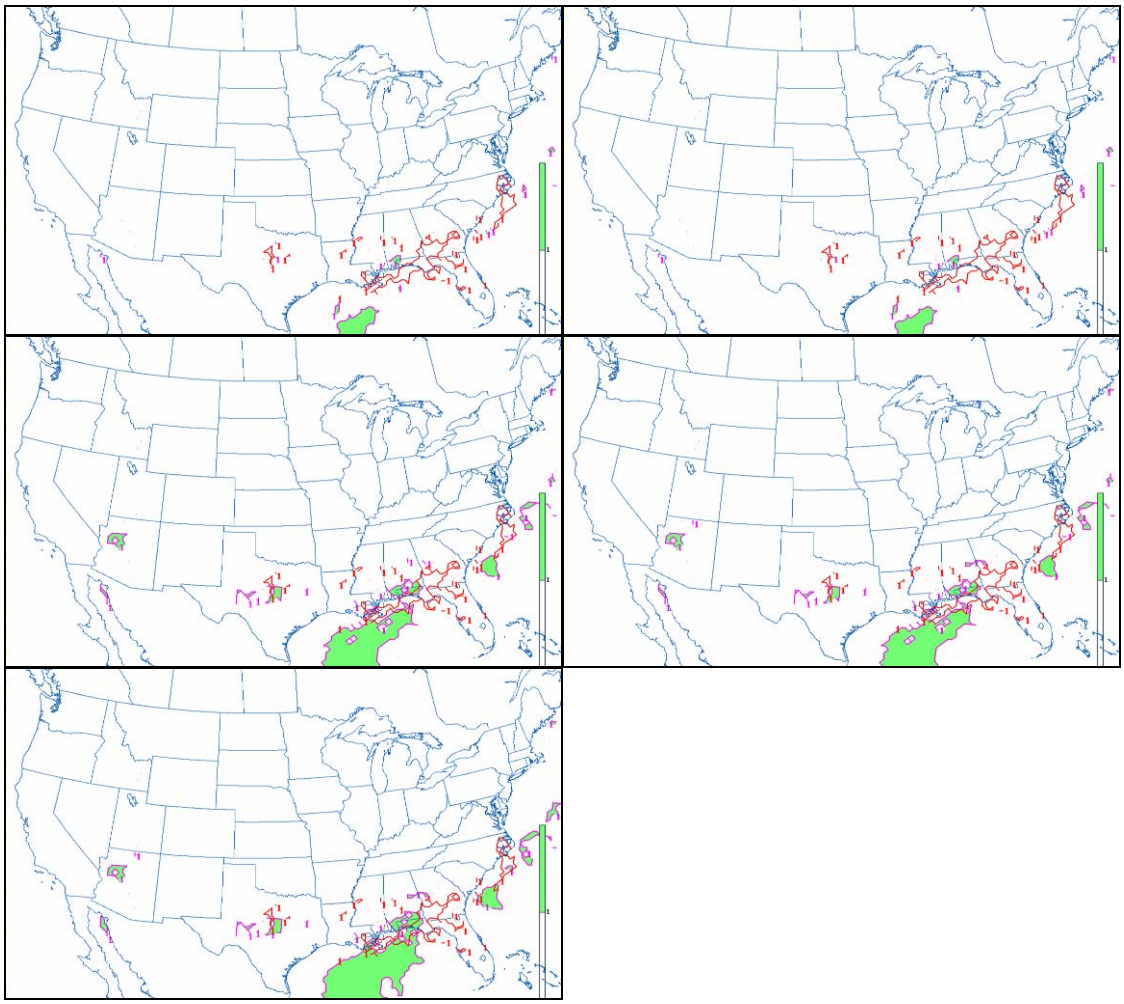

Figure 7. Configuration group B output for NAM 01 Feb 2007 0000Z run, 18-hour forecast (green shade, purple outline) and observed severe weather at 1800Z, 01 Feb 2007 (red outline). Configurations are ordered as follows: B1 (upper left), B2 (upper right), B3 (middle left), B4 (middle right), and B5 (lower left).

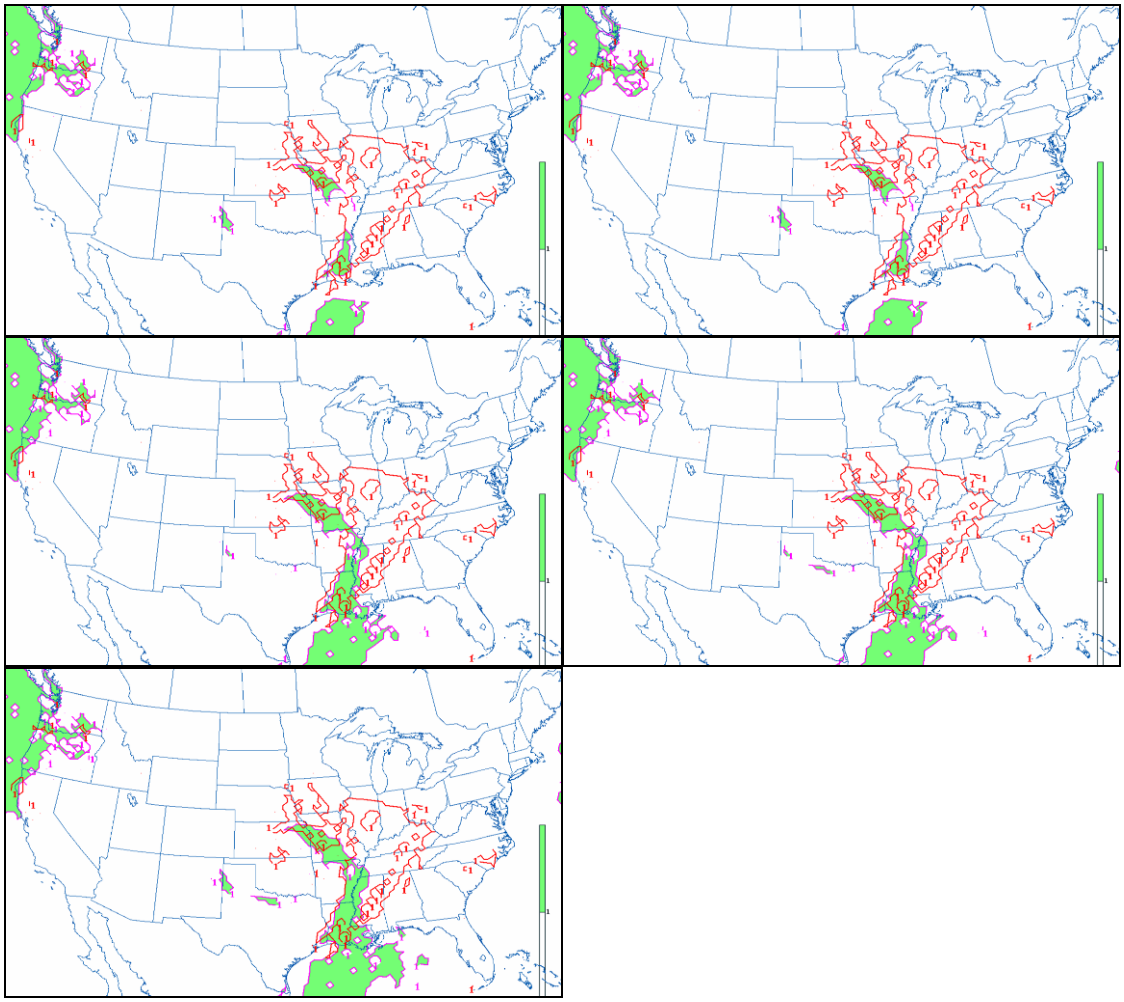

Figure 8. Configuration group B output for NAM 24 Feb 2007 0000Z run, 24-hour forecast (green shade, purple outline) and observed severe weather at 0000Z, 25 Feb 2007 (red outline). Configurations are ordered as follows: B1 (upper left), B2 (upper right), B3 (middle left), B4 (middle right), and B5 (lower left).

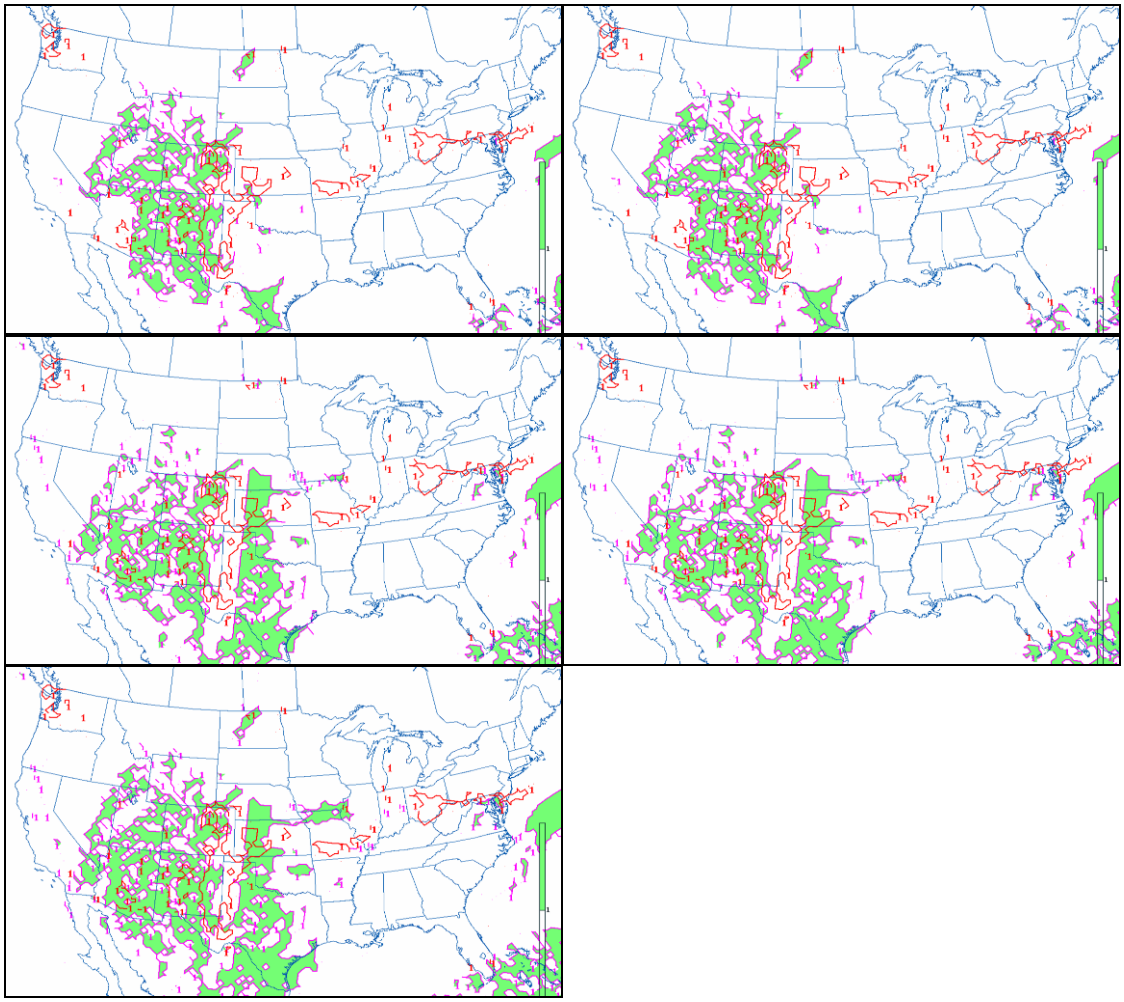

Figure 9. Configuration group B output for NAM 23 Mar 2007 1200Z run, 12-hour forecast (green shade, purple outline) and observed severe weather at 0000Z, 24 Mar 2007 (red outline). Configurations are ordered as follows: B1 (upper left), B2 (upper right), B3 (middle left), B4 (middle right), and B5 (lower left).

3. Configuration Group C

Tier I performance data for algorithm configurations in Group C are listed in the following four tables. The forecast solutions were constrained by requiring the K-Index to be above its minimum threshold in addition to one or more of the remaining three indices meeting its respective threshold for the algorithm to forecast instability. For the NAM 00Z, 01 Feb 2007, 18-hour forecast (Table 15 and Figure 10), configuration C6 demonstrated the best performance in terms of ROC and KSS (0.5770 and 0.509, respectively). This configuration also outperformed the remaining five configurations in forecast accuracy at 87.6%. The configuration with the second-highest performance statistics was configuration C5, with a ROC of 0.5320, KSS of 0.447, and accuracy of 86.7%.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
C1	4.2%	81.8%	0.229	0.035	-0.010	0.010	0.1120	83.4%
C2	27.1%	58.1%	0.646	0.197	0.102	0.207	0.3450	84.0%
C3	41.7%	50.0%	0.833	0.294	0.214	0.346	0.4585	85.5%
C4	41.7%	50.0%	0.833	0.294	0.214	0.346	0.4585	85.5%
C5	52.1%	45.7%	0.958	0.363	0.291	0.447	0.5320	86.7%
C6	58.3%	42.9%	1.021	0.406	0.339	0.509	0.5770	87.6%

Table 15. Tier I experiment results using the 18-hour forecast from the 01 Feb 2007 0000Z NAM run for configuration group C.

Performance data for configuration group C for the NAM 00Z, 24 Feb 07, 24-hour forecast is listed in Table 16, and graphically displayed in Figure 11. Overall, performance results for this case were very close among configurations C2 through C6. Configurations C3 and C4 were exactly tied with a ROC of 0.6105 and KSS of 0.458. Configuration C6 narrowly outperformed C5 with a ROC of 0.6285 versus 0.6270, and KSS of 0.353 versus 0.348, respectively. Considering the ±5% allowable margin of error, configurations C5 and C6 ended up in a statistical tie for this case. Accuracy statistics largely followed KSS and ROC since there were a large number of gridboxes with observed severe thunderstorms, relative to the other two cases studies. Therefore,

accuracy reflects a robust sample of both correctly-forecasted non-occurrences and correctly forecasted hits for severe thunderstorms. As for the previous configuration groups, the verification data shows a large region of "hits" over the Ohio Valley, which is believed to be heavy stratoform precipitation due to overrunning. An attempt to exclude this region from the performance statistics was made by discarding any gridpoints where the near-surface temperature was below 18°C in accordance with rules established in AWS TR-200 (Miller 1972).

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
C1	21.2%	27.4%	0.274	0.204	0.049	0.204	0.4690	85.5%
C2	38.6%	25.0%	0.515	0.342	0.223	0.359	0.5680	87.2%
C3	49.8%	27.7%	0.689	0.418	0.320	0.458	0.6105	88.0%
C4	49.8%	27.7%	0.689	0.418	0.320	0.458	0.6105	88.0%
C5	53.1%	27.7%	0.734	0.441	0.348	0.489	0.6270	88.4%
C6	53.9%	28.2%	0.751	0.445	0.353	0.495	0.6285	88.4%

Table 16. Tier I experiment results using the 24-hour forecast from the 24 Feb 2007 0000Z NAM run for configuration group C.

None of the configurations in group C forecasted severe thunderstorms over the region along the Texas/New Mexico border northward into western Kansas and eastern Colorado, where severe thunderstorms were observed either directly or indirectly via radar imagery at 00Z, 24 Mar 2007. Meanwhile, all configurations indicated favorable conditions for severe thunderstorms over central Texas, where no severe thunderstorms were recorded. The result for this case was POD values in the single-digits, FAR rates between 95% and 100%, negative KSS for all configurations, and ROC areas ranging from zero to 0.0355. The large number of correctly-forecasted non-occurrences led to the accuracy rate being in the 80%-90% range and does not reflect on the failure of this configuration group to reliably forecast the location where severe thunderstorms occurred during this particular event. Results for the 00Z 24 Mar 07 case are presented statistically in Table 17 and graphically in Figure 12.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
C1	0.0%	100.0%	0.154	0.000	-0.076	-0.014	0.0000	90.6%
C2	1.0%	97.4%	0.386	0.007	-0.055	-0.023	0.0180	88.8%
C3	2.4%	96.9%	0.779	0.014	-0.034	-0.043	0.0275	85.8%
C4	2.4%	96.9%	0.802	0.013	-0.033	-0.045	0.0275	85.6%
C5	3.0%	97.1%	1.044	0.015	-0.026	-0.060	0.0295	83.8%
C6	4.2%	97.1%	1.489	0.017	-0.016	-0.086	0.0355	80.3%

Table 17. Tier I experiment results using the 12-hour forecast from the 23 Mar 2007 1200Z NAM run for configuration group C.

The overall results for configuration group C suggest that requiring the K-Index to meet its minimum threshold in addition to one more of the remaining stability indices placed too much constraint on the forecast solution. Therefore, the algorithm generally underforecasted the size of the region favored for severe thunderstorm development. The underforecasting contributed to the performance statistics ending up being substantially lower than those for configuration groups A and B. Because configuration C6 has the lowest threshold settings, and thus the least amount of constraint, it performed better than the remaining configurations in group C, and was therefore retained for Tier II analysis.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
C1	8.5%	69.7%	0.219	0.080	-0.012	0.067	0.1940	86.5%
C2	22.2%	60.2%	0.515	0.182	0.090	0.181	0.3100	86.7%
C3	31.3%	58.2%	0.767	0.242	0.167	0.254	0.3655	86.4%
C4	28.7%	58.2%	0.775	0.242	0.167	0.253	0.3525	86.3%
C5	36.1%	56.8%	0.760	0.273	0.204	0.292	0.3965	86.3%
C6	38.8%	56.1%	1.087	0.289	0.225	0.306	0.4135	85.4%

Table 18. Overall Tier I performance for configuration group C.

Forecast and verification data for configuration group C is displayed in the following figures. The green-shaded regions indicate areas where favorable conditions

for severe thunderstorms were indicated by the 3E Algorithm. The areas where severe thunderstorms were reported or where radar reflectivity values exceeded 40 dBZ are outlined in light blue.

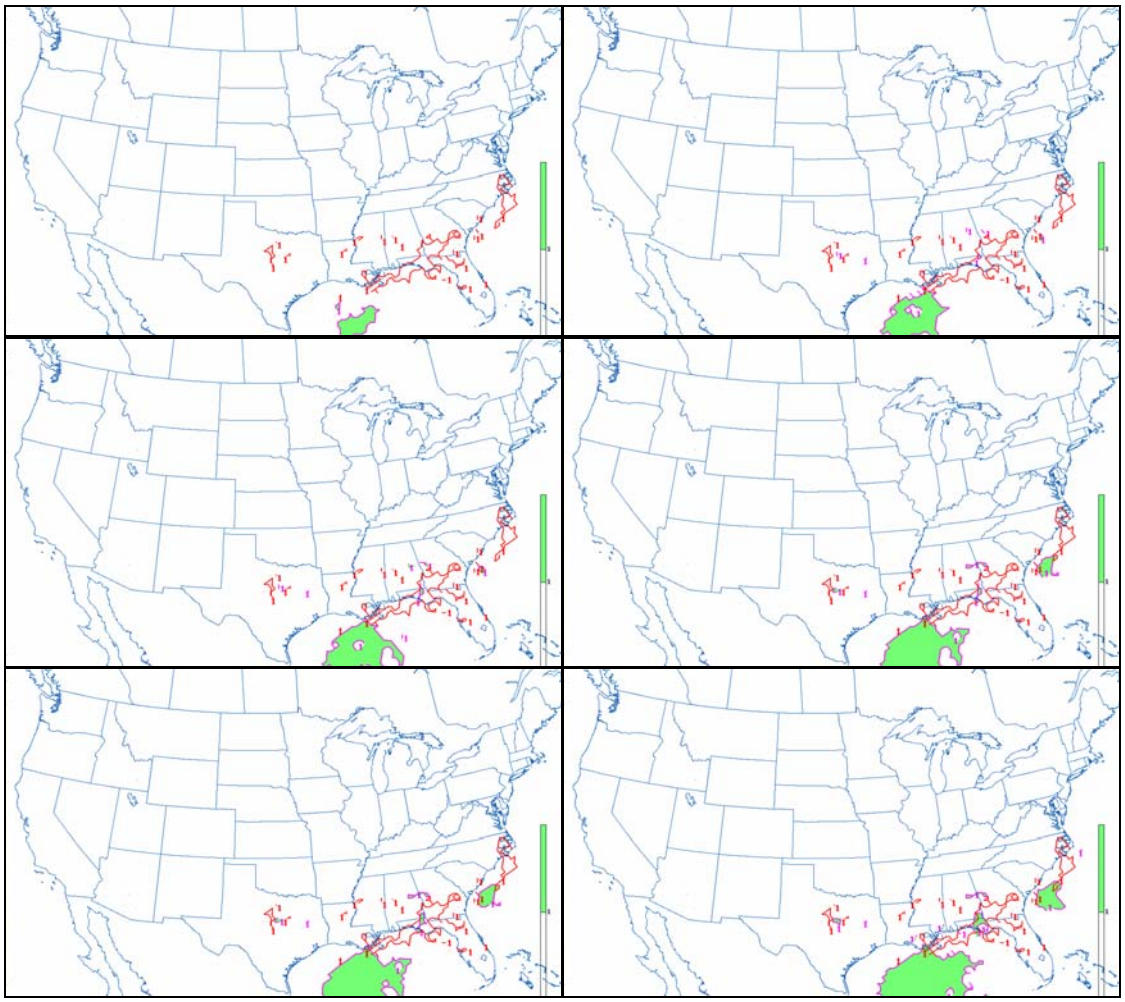

Figure 10. Configuration group C output for NAM 01 Feb 2007 0000Z run, 18-hour forecast (green shade, purple outline) and observed severe weather at 1800Z, 01 Feb 2007 (red outline). Configurations are ordered as follows: C1 (upper left), C2 (upper right), C3 (middle left), C4 (middle right), C5 (lower left), and C6 (lower right).

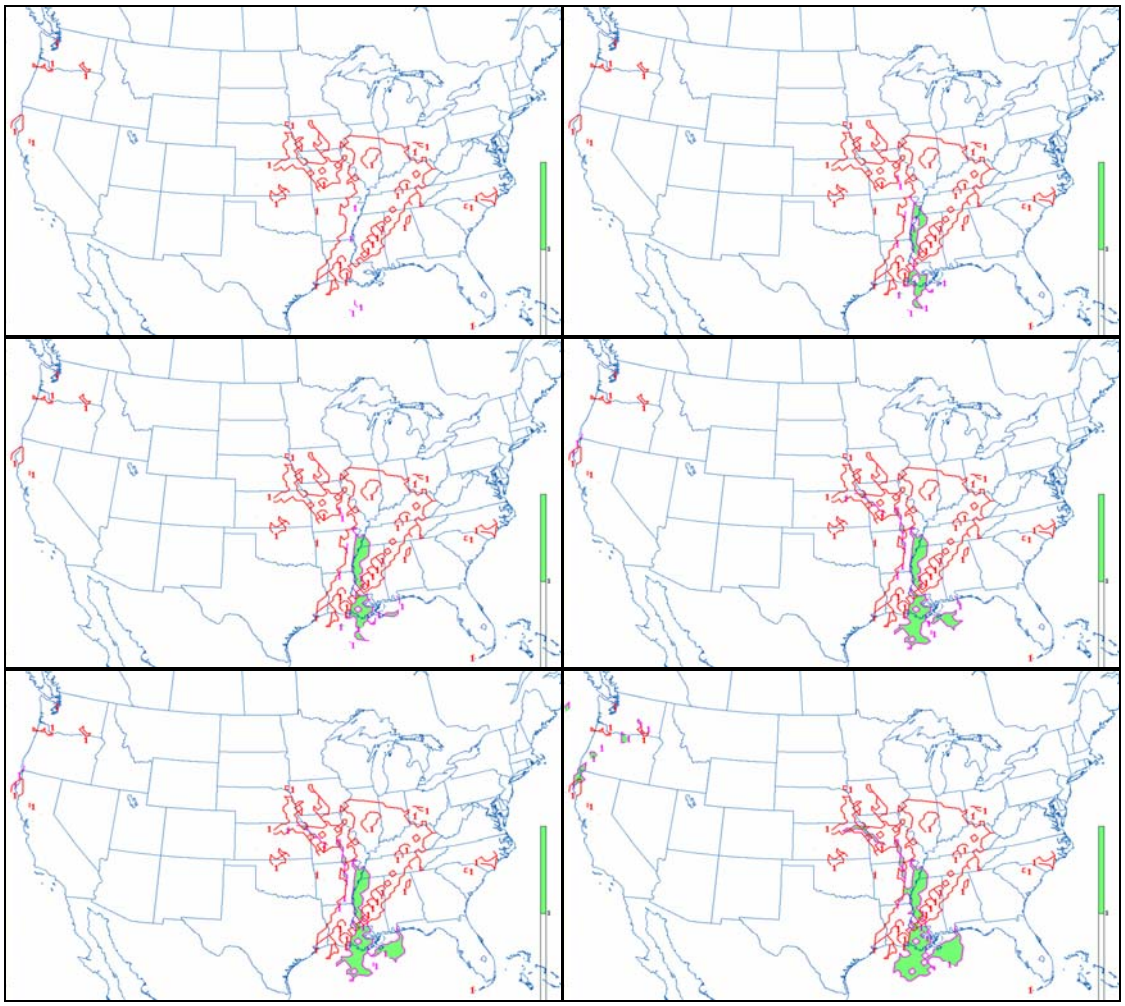

Figure 11. Configuration group C output for NAM 24 Feb 2007 0000Z run, 24-hour forecast (green shade, purple outline) and observed severe weather at 0000Z, 25 Feb 2007 (red outline). Configurations are ordered as follows: C1 (upper left), C2 (upper right), C3 (middle left), C4 (middle right), C5 (lower left), and C6 (lower right).

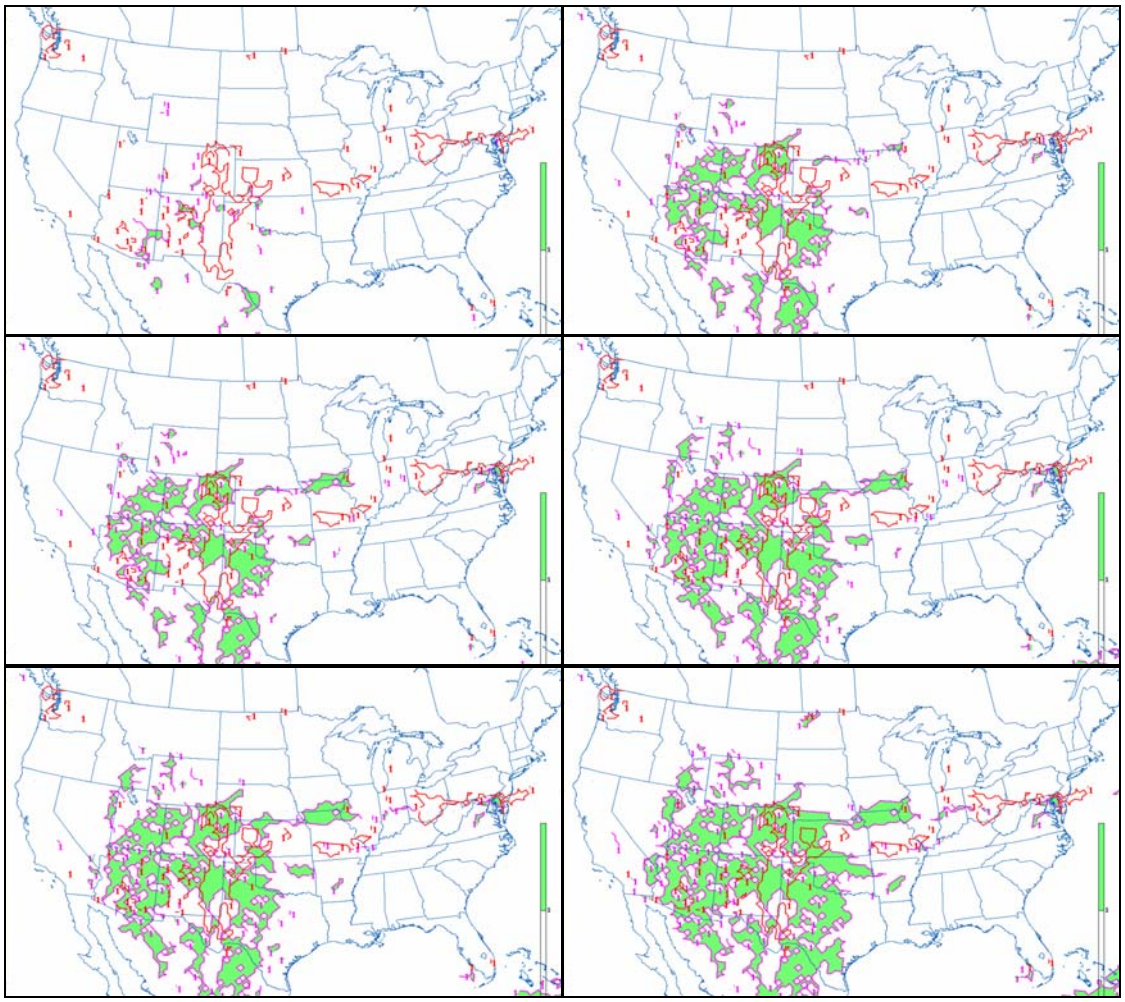

Figure 12. Configuration group C output for NAM 23 Mar 2007 1200Z run, 12-hour forecast (green shade, purple outline) and observed severe weather at 0000Z, 24 Mar 2007 (red outline). Configurations are ordered as follows: C1 (upper left), C2 (upper right), C3 (middle left), C4 (middle right), C5 (lower left), and C6 (lower right).

4. Configuration Group D

Configuration group D included the original set of AFWA proxies plus the horizontal CAPE gradient as an additional dynamic forcing term. Because CAPE is a function of the surface temperature, moisture content, and environmental lapse rate, a strong CAPE gradient can be a good indicator to the presence of fronts, drylines, convergence boundaries, all of which are good forcing mechanisms for severe thunderstorms initiation. Starting with the horizontal CAPE gradient at 1 J kg^{-1} km^{-1} and keeping the remaining proxies at AFWA defaults (configuration D1) yielded results nearly identical to the results of configurations A1, B1, and C1 for the NAM 00Z, 01 Feb 2007, 18-hour forecast. Thresholds were lowered for Lifted Index, CAPE, and CIN for configurations D2 through D4, and lessening the constraint on the forecast solution. The increase in aerial coverage for severe thunderstorms forecasted by the algorithm is again the result of the reduced threshold settings. Ultimately that led to an increase in the verification of severe thunderstorm occurrences (Table 19 and Figure 13). Configuration D4 has the lowest threshold settings within this group; its results included a ROC area of 0.5960 and KSS of 0.544. This compares with a 0.4635 ROC area and KSS of 0.356 for the second-ranked configuration (D3) in this group.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
D1	4.2%	83.3%	0.250	0.034	-0.097	0.006	0.1045	83.1%
D2	29.2%	58.8%	0.708	0.206	0.115	0.221	0.3520	83.7%
D3	43.8%	51.1%	0.896	0.300	0.223	0.356	0.4635	85.2%
D4	62.5%	43.3%	1.104	0.423	0.360	0.544	0.5960	87.6%

Table 19. Tier I experiment results using the 18-hour forecast from the 01 Feb 2007 0000Z NAM run for configuration group D.

Configuration D4 had the highest ROC area (0.6545) and KSS (0.578) for the NAM 00Z, 24 Feb 2007, 24-hour forecast. The ROC area and KSS differences between configurations D3 and D4 were both within the ±5% margin of uncertainty (0.9% and 3.6%, respectively). Therefore a statistical tie resulted between configurations D3 and D4 for this case. The ROC area difference between configurations D2 and D4 was also

within tolerance (4.7%), but configuration D4 clearly outperformed D2 in terms of KSS (15.2%), thus no statistical tie existed between configurations D2 and D4. The results for each configuration are depicted in Figure 14. Again it is noted how the least-constrained solution (D4) had the largest aerial coverage in its forecast for severe thunderstorms, and also had the largest ROC area and KSS.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
D1	41.3%	31.0%	0.602	0.350	0.239	0.376	0.5515	86.7%
D2	53.5%	28.7%	0.751	0.440	0.348	0.490	0.6240	88.3%
D3	61.8%	32.2%	0.913	0.478	0.397	0.557	0.6480	88.3%
D4	64.7%	33.8%	0.979	0.486	0.410	0.578	0.6545	88.2%

Table 20. Tier I experiment results using the 24-hour forecast from the 24 Feb 2007 0000Z NAM run for configuration group D.

By including the horizontal CAPE gradient as an additional dynamic forcing proxy, configuration group D increased the POD from the previous configuration groups for the NAM 12Z, 23 Mar 2007, 12-hour forecast. However, this was only accomplished through gross overforecasting of the severe thunderstorm potential, which resulted in the FAR ranging between 83% and 87% and ROC areas for all configurations falling below the "no-skill" threshold for this case. Figure 15 depicts how each of the configurations in group D generally saturate the model domain over the southern Rockies in an attempt to encompass the area where severe thunderstorms occurred, while also retaining the region over central and southern Texas in their forecasts for severe weather.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
D1	73.9%	83.3%	4.616	0.152	0.137	0.394	0.4530	66.1%
D2	74.9%	84.4%	4.800	0.149	0.134	0.389	0.4525	64.8%
D3	76.0%	85.1%	5.109	0.142	0.129	0.372	0.4545	62.4%
D4	77.6%	86.1%	5.600	0.133	0.121	0.346	0.4575	58.7%

Table 21. Tier I experiment results using the 12-hour forecast from the 23 Mar 2007 1200Z NAM run for configuration group D.

Configuration D4 clearly outperformed the other three configurations within group D, and was therefore selected to be run on the entire dataset across all six cases. Its overall ROC area of 0.5695 and KSS of 0.489 placed it well ahead of the next-ranked setting. Configuration D3 had a KSS of 0.428 and a ROC area of 0.5215, which beats the "no skill" ROC area of 0.5, but lags configuration D4 by 0.048 (8.4%) in ROC area, and by 0.061 (12.5%) in KSS.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
D1	39.9%	66.1%	1.823	0.179	0.093	0.259	0.3690	78.6%
D2	52.6%	57.3%	2.086	0.265	0.200	0.368	0.4765	78.9%
D3	60.5%	56.2%	2.306	0.307	0.250	0.428	0.5215	78.6%
D4	68.3%	54.4%	2.561	0.347	0.293	0.489	0.5695	78.2%

Table 22. Overall Tier I performance for configuration group D.

Graphical algorithm forecasts overlaid with Storm Reports and radar verification data are presented in Figures 13-15. Note the tendency for overforecasting in all configurations and all three cases.

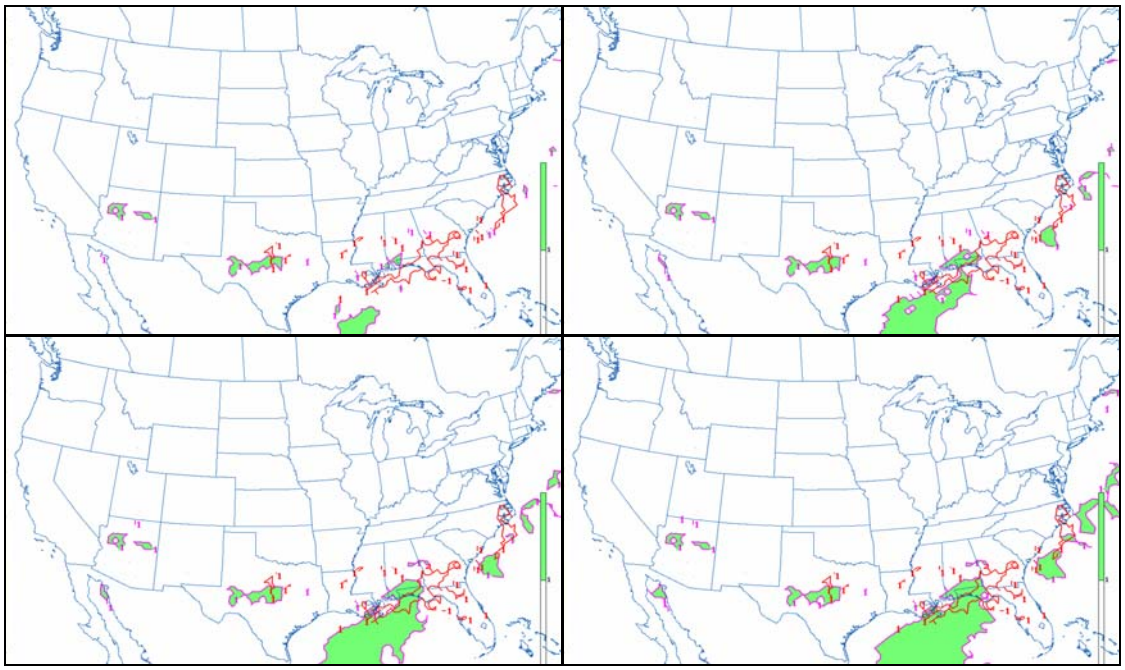

Figure 13. Configuration group D output for NAM 01 Feb 2007 0000Z run, 18-hour forecast (green shade, purple outline) and observed severe weather at 1800Z, 01 Feb 2007 (red outline). Configurations are ordered as follows: D1 (upper left), D2 (upper right), D3 (lower left), and D4 (lower right).

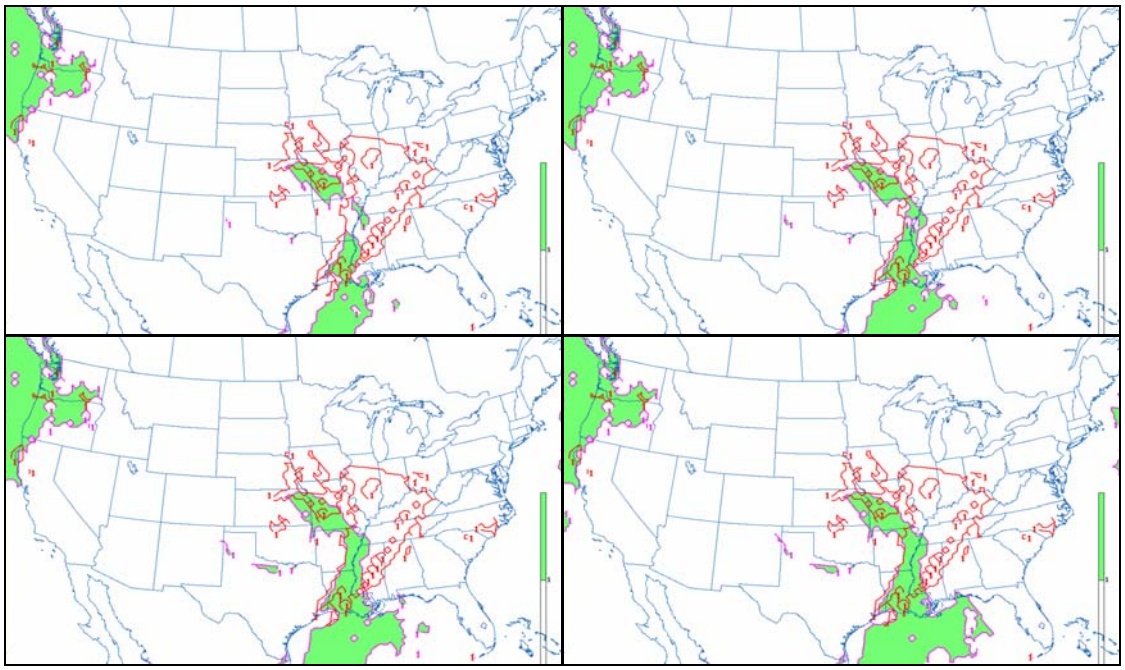

Figure 14. Configuration group D output for NAM 24 Feb 2007 0000Z run, 24-hour forecast (green shade, purple outline) and observed severe weather at 0000Z, 25 Feb 2007 (red outline). Configurations are ordered as follows: D1 (upper left), D2 (upper right), D3 (lower left), and D4 (lower right).

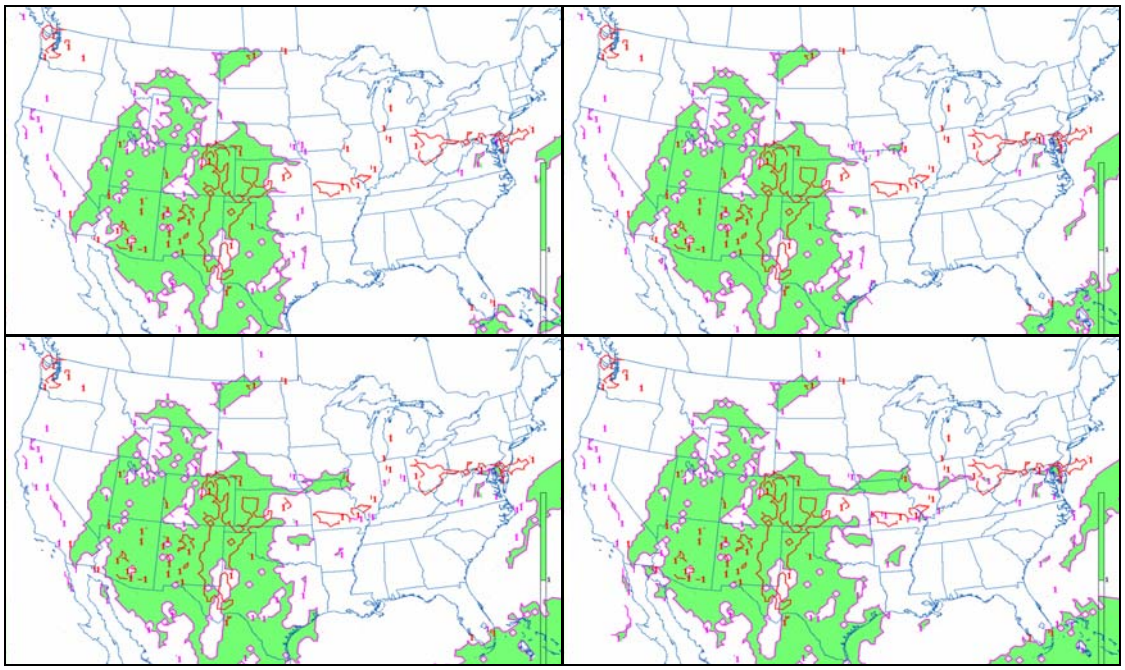

Figure 15. Configuration group D output for NAM 23 Mar 2007 1200Z run, 12-hour forecast (green shade, purple outline) and observed severe weather at 0000Z, 24 Mar 2007 (red outline). Configurations are ordered as follows: D1 (upper left), D2 (upper right), D3 (lower left), and D4 (lower right).

5. Configuration Group E

Configuration group E appears to be identical to group D by again considering the horizontal CAPE gradient. However, the main difference between groups D and E is that configuration group E required the CAPE gradient to meet its threshold for dynamic forcing to be present. Configuration group D did not, as long as one or more of the remaining proxies met its respective threshold. The following tables contain performance data for configuration group E. Configuration E4 had the highest POD and lowest FAR for the NAM 00Z, 01 Feb 2007, 18-hour forecast. As such it also had the greatest ROC area (0.5960) and highest KSS (0.544). Performance of the remaining configurations fell well below that of E4, with the latter outperforming configuration E3 by 23.1% in ROC area and KSS by 36.4%. Results for the first case are numerically listed in Table 23, and presented graphically in Figure 16.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
E1	4.2%	81.8%	0.229	0.035	-0.099	0.010	0.1120	83.4%
E2	27.1%	58.1%	0.646	0.197	0.102	0.207	0.3450	84.0%
E3	41.7%	50.0%	0.833	0.294	0.214	0.346	0.4585	85.5%
E4	62.5%	43.3%	1.104	0.423	0.360	0.544	0.5960	87.6%

Table 23. Tier I experiment results using the 18-hour forecast from the 01 February 2007 0000Z NAM run for configuration group E.

Results from the NAM 00Z, 24 Feb 2007, 24-hour forecast reflected a fairly narrow spread in the solutions generated by configurations E2, E3, and E4, with E4 having the best handle on the forecast. Except for configuration E1, settings in group E resulted in ROC areas that generally exceeded the "no-skill" ROC area by a margin of 17%-23%. The greatest increase in ROC area occurred with configuration E4 (0.6545). KSS outputs generally followed the ROC trends with each configuration. Configuration E4 had the highest KSS at 0.578, as opposed to 0.188, 0.452, and 0.511 for configurations E1, E2, and E3, respectively. Statistics for configuration group E are listed in Table 24. Graphical outputs are displayed in Figure 17.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
E1	20.7%	31.5%	0.303	0.189	0.038	0.188	0.4460	84.7%
E2	49.4%	28.7%	0.693	0.412	0.312	0.452	0.6035	87.8%
E3	56.8%	32.8%	0.846	0.445	0.358	0.511	0.6200	87.7%
E4	64.7%	33.8%	0.979	0.486	0.410	0.578	0.6545	87.8%

Table 24. Tier I experiment results using the 24-hour forecast from the 24 February 2007 0000Z NAM run for configuration group E.

Results from the NAM 12Z, 23 Mar 2007, 12-hour forecast indicated that configurations in group E possessed low scores in both KSS and ROC due to the FAR ranging from 81% to 86% across all four settings. Lowering the index thresholds enabled more areas affected by severe thunderstorms to be included in the forecast solution. However, it also expanded the region over which the forecasted severe thunderstorms, thereby increasing the FAR (Figure 18). Because the POD increased at a much faster rate for each setting (65.1 percentage point increase from E1 to E4), than the FAR (5.1 percentage point increase from E1 to E4), the result was a net improvement in the KSS and ROC area of configurations E2, E3, and E4, relative to configuration E1 and the default setting.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
E1	12.5%	81.0%	0.661	0.082	0.030	0.078	0.1575	88.5%
E2	62.7%	83.6%	3.820	0.149	0.132	0.342	0.3955	70.8%
E3	64.6%	84.5%	4.117	0.142	0.126	0.326	0.4005	68.5%
E4	77.6%	86.1%	5.600	0.133	0.121	0.346	0.4575	66.4%

Table 25. Tier I experiment results using the 12-hour forecast from the 23 March 2007 1200Z NAM run for configuration group E.

Requiring the horizontal CAPE gradient to meet its specified threshold in addition to one or more of the remaining dynamic forcing proxies meeting its respective threshold constrained the forecast solution, but this constraint had little impact on the algorithm's overall performance in comparison to configuration group D. In fact, the results suggest that while the constraints introduced in configuration group E led to a slight (1-2

percentage points) decrease in the FAR from configuration group D, the same constraints also reduced the POD by a much greater amount (6 to 27 percentage points) from configuration group D. Therefore, the results from configuration group E reflected an overall degradation of performance from configuration group D. Configuration E4 was the only configuration within this group to have a ROC area (0.5695) that exceeded the "no-skill" ROC area of 0.5. In terms of ROC area and KSS, configuration E4 surpassed the next-best configuration (E3) by 13.8% and 19.4%, respectively.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
E1	12.5%	64.8%	0.398	0.102	-0.010	0.092	0.2385	85.5%
E2	46.4%	58.9%	1.720	0.253	0.182	0.334	0.4375	80.9%
E3	54.0%	55.8%	1.932	0.294	0.233	0.394	0.4910	80.6%
E4	68.3%	54.4%	2.561	0.347	0.293	0.489	0.5695	80.6%

Table 26. Overall Tier I performance for configuration group E.

The figures on the following pages depict the forecasted locations of severe thunderstorms for each of the three cases examined and four settings within configuration group E. Requiring the CAPE gradient to meet its threshold added constraint to the forecast solutions which mitigated the amount of overforecasting, compared to configuration group D, but it also reduced the area where correctly-forecasted events occurred, thereby increasing the miss rate. Nonetheless, configuration group E had less constraint on the solution than configuration groups A and C. Therefore, severe thunderstorms forecasts from group typically encompassed a larger area than the two aforementioned groups, leading to increased chances of verification.

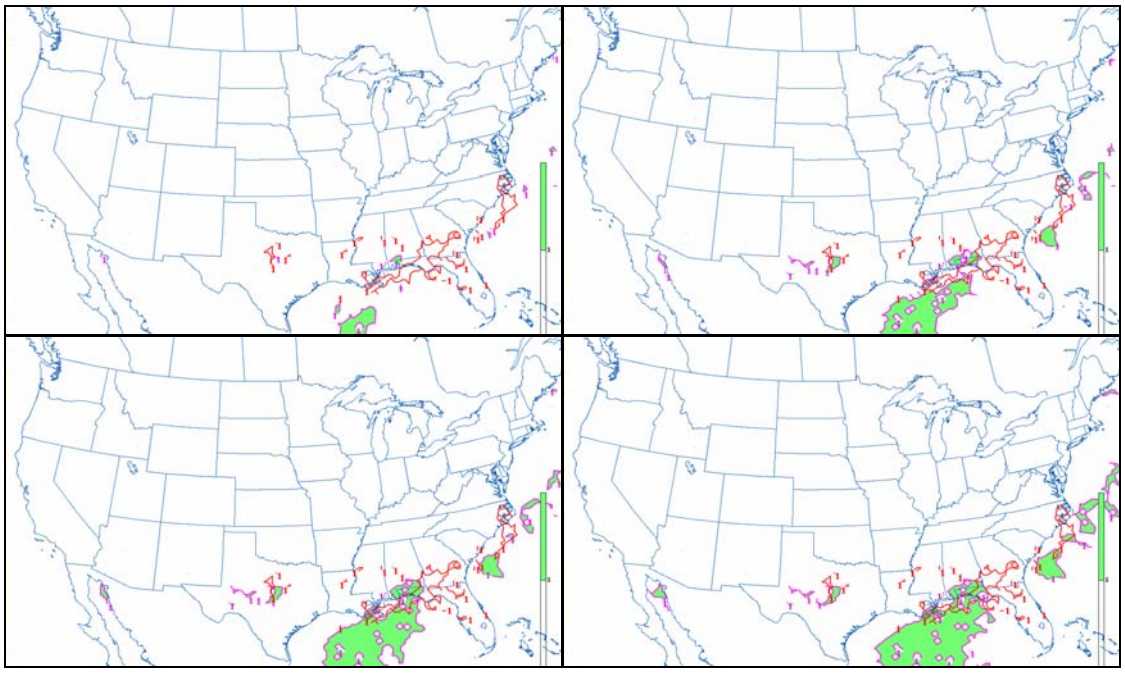

Figure 16. Configuration group E output for NAM 01 Feb 2007 0000Z run, 18-hour forecast (green shade, purple outline) and observed severe weather at 1800Z, 01 Feb 2007 (red outline). Configurations are ordered as follows: E1 (upper left), E2 (upper right), E3 (lower left), and E4 (lower right).

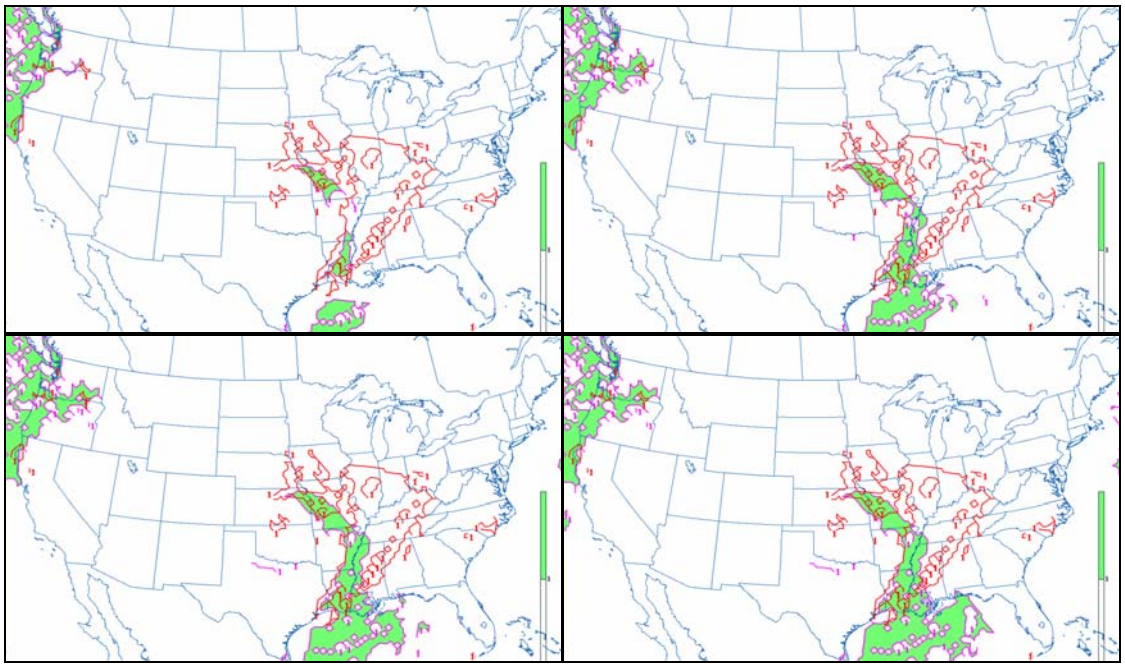

Figure 17. Configuration group E output for NAM 24 Feb 2007 0000Z run, 24-hour forecast (green shade, purple outline) and observed severe weather at 0000Z, 25 Feb 2007 (red outline). Configurations are ordered as follows: E1 (upper left), E2 (upper right), E3 (lower left), and E4 (lower right).

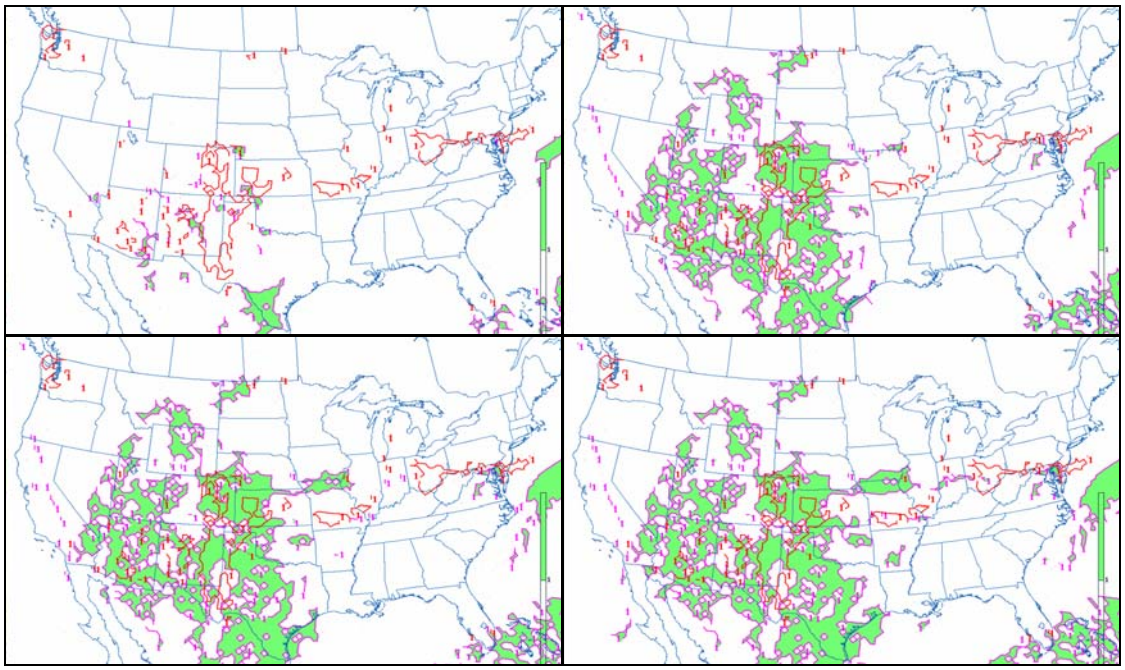

Figure 18. Configuration group E output for NAM 23 Mar 2007 1200Z run, 12-hour forecast (green shade, purple outline) and observed severe weather at 0000Z, 24 Mar 2007 (red outline). Configurations are ordered as follows: E1 (upper left), E2 (upper right), E3 (lower left), and E4 (lower right).

6. Configuration Group F

Configuration group F was unique from the other configuration groups because an attempt was made to examine the algorithm's performance when multiple modifications to the algorithm were implemented simultaneously. Adding in the modified instability and forcing proxies over the Intermountain West along with the horizontal CAPE gradient over the remainder of the CONUS (configurations F1 through F3) yielded little change from the algorithm output using the original set of proxies for the NAM 00Z, 01 Feb 2007, 18-hur forecast. The ROC areas of 0.1045, 0.3520, and 0.4630 for configurations F1, F2, and F3, respectively, reflected this. The most substantial improvement occurred when the K-Index was included in addition to the high-elevation proxies and CAPE gradient (configurations F4 and F5). Both settings ended up with ROC areas that exceeded the "no-skill" threshold of 0.500, with the ROC area of configuration F4 being 0.5565, and 0.5420 for configuration F5. Consistent with these findings, the KSS for configurations F4 and F5 were 0.494 and 0.465, respectively, compared to 0.006 for configuration F1, 0.121 for F2, and 0.360 for F3. Figure 19 depicts how the region favored for severe thunderstorms increased in size due to the cumulative effects of all of the modifications being added to the algorithm simultaneously.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
F1	4.2%	83.3%	0.250	0.034	-0.097	0.006	0.1045	83.1%
F2	29.2%	58.8%	0.708	0.206	0.115	0.121	0.3520	83.7%
F3	43.8%	51.2%	0.896	0.300	0.223	0.360	0.4630	85.2%
F4	58.5%	47.2%	1.104	0.384	0.319	0.494	0.5565	86.4%
F5	54.2%	45.8%	1.000	0.371	0.302	0.465	0.5420	86.7%

Table 27. Tier I experiment results using the 18-hour forecast from the 01 February 2007 0000Z NAM run for configuration group F.

The ROC areas for configurations F2 through F5 exceeded the "no-skill" ROC area of 0.5, for the NAM 00Z, 24 Feb 2007, 24-hour forecast. KSS results were consistent with the ROC areas, ranging from 0.480 for configuration F2 to 0.708 for

configuration F5. The KSS for configuration F5 was slightly lower than F4 at 0.645 since the more-constrained solution for F5 reduced the number of gridpoints would have otherwise been correctly forecasted severe thunderstorm events. Although this constraint also cut the FAR, the POD decreased 7.1 percentage points between F4 and F5, while the FAR decreased 1.3 percentage points. That led to configuration F5 having a somewhat lower forecast skill than configuration F4. The results for each configuration are graphically presented in Figure 20.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
F1	22.8%	40.9%	0.386	0.197	0.057	0.195	0.4095	84.0%
F2	53.5%	33.2%	0.801	0.423	0.332	0.480	0.6015	87.4%
F3	61.8%	35.7%	0.963	0.460	0.388	0.546	0.6305	87.5%
F4	83.0%	41.2%	1.411	0.525	0.467	0.708	0.7090	87.0%
F5	75.9%	42.5%	1.320	0.487	0.423	0.645	0.6670	86.2%

Table 28. Tier I experiment results using the 24-hour forecast from the 24 February 2007 0000Z NAM run for configuration group F.

Like the previous configuration groups, FAR values for group F ranged in the 84%-88% range for the NAM 12Z, 23 Mar 2007, 12-hour forecast. ROC areas ranged from 0.2655 to 0.3675. POD values varied from 37.4% for configuration F1 to 61.0% for configuration F4. Graphically, Figure 21 depicts how the more constrained solution for configuration F5 resulted in a smaller area for severe thunderstorms forecasted by the algorithm compared to configuration F4.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
F1	37.4%	84.3%	2.379	0.124	0.100	0.195	0.2655	78.5%
F2	41.6%	87.7%	3.387	0.105	0.086	0.151	0.2695	70.9%
F3	42.6%	88.5%	3.719	0.099	0.082	0.133	0.2705	68.4%
F4	61.0%	87.2%	4.764	0.118	0.104	0.239	0.3690	62.8%
F5	60.0%	86.5%	4.465	0.124	0.108	0.257	0.3675	65.1%

Table 29. Tier I experiment results using the 12-hour forecast from the 23 March 2007 1200Z NAM run for configuration group F.

Configuration F4 included the horizontal CAPE gradient as a forcing proxy, but did not specifically require it to meet its threshold as long as one or more of the remaining forcing proxies had reached its respective threshold. Configuration F5 did require the horizontal CAPE gradient to meet its threshold in addition to one or more of the remaining proxies reaching its threshold for the algorithm to indicate dynamic forcing in the forecast. Because the solution for configuration F5 was more constrained than F4, the results indicated that configuration F4 narrowly outperformed configuration F5.

Config	POD	FAR	Bias	TS	ETS	KSS	ROC	ACC
F1	21.5%	69.5%	1.005	0.118	0.020	0.132	0.2600	81.9%
F2	41.4%	59.9%	1.632	0.245	0.178	0.284	0.4075	80.7%
F3	51.2%	58.5%	1.859	0.286	0.228	0.346	0.4635	80.4%
F4	67.4%	58.5%	2.426	0.342	0.297	0.480	0.5445	78.7%
F5	63.4%	58.2%	2.262	0.327	0.278	0.457	0.5260	79.3%

Table 30. Overall Tier I performance for configuration group F.

Algorithm forecast solutions for configuration group F, with SPC Storm Reports and WSR-88D overlaid for verification, are presented in the figures on the following pages. .Configurations in group F tended to have the largest aerial coverage in their severe thunderstorm forecasts, compared to configurations in other groups. While members of configuration group F tended to overforecast severe thunderstorms over the Intermountain West, this overforecasting was not as extreme as the complete saturation of the grids as observed with configurations in group D.

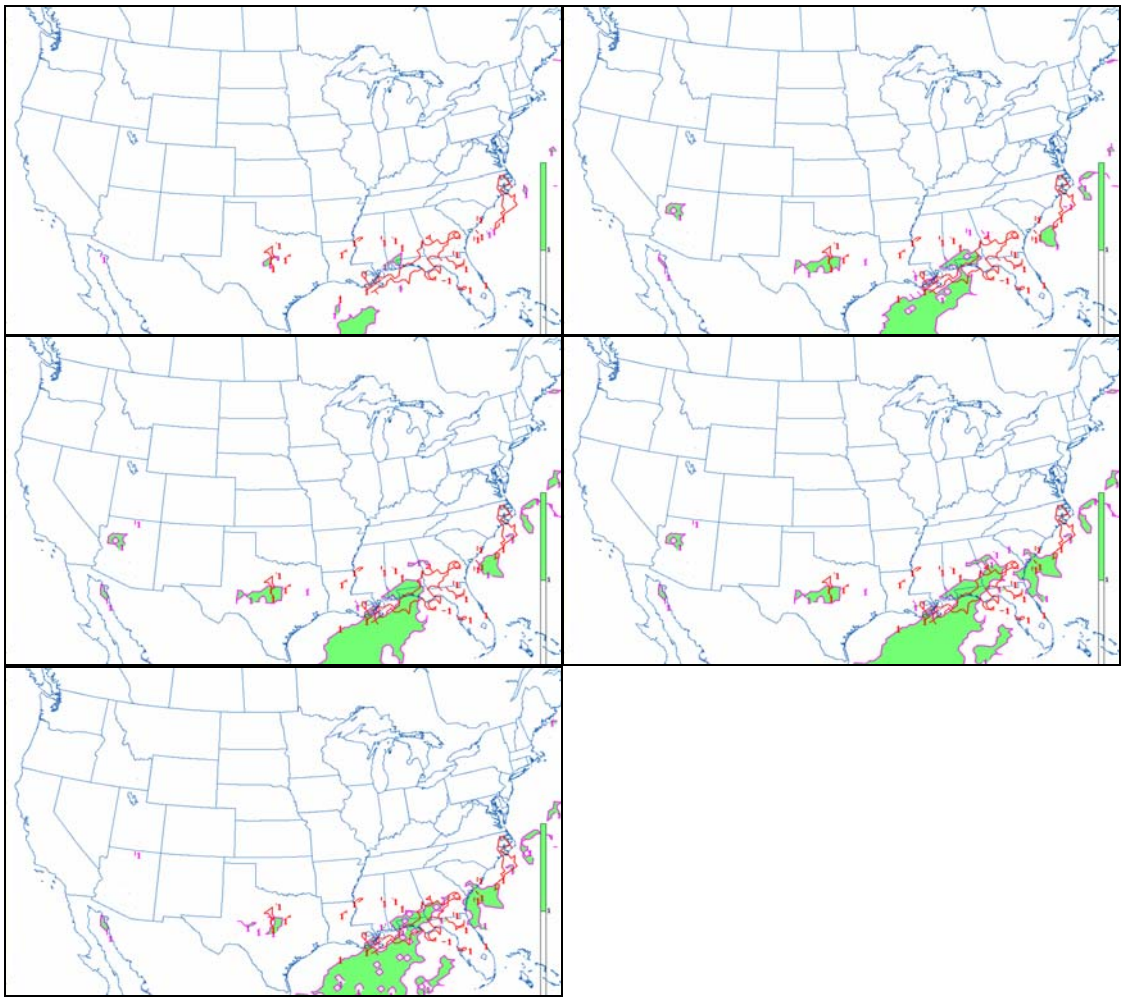

Figure 19. Configuration group F output for NAM 01 Feb 2007 0000Z run, 18-hour forecast (green shade, purple outline) and observed severe weather at 1800Z, 01 February 2007 (red outline). Configurations are ordered as follows: F1 (upper left), F2 (upper right), F3 (middle left), F4 (middle right), and B5 (lower left).

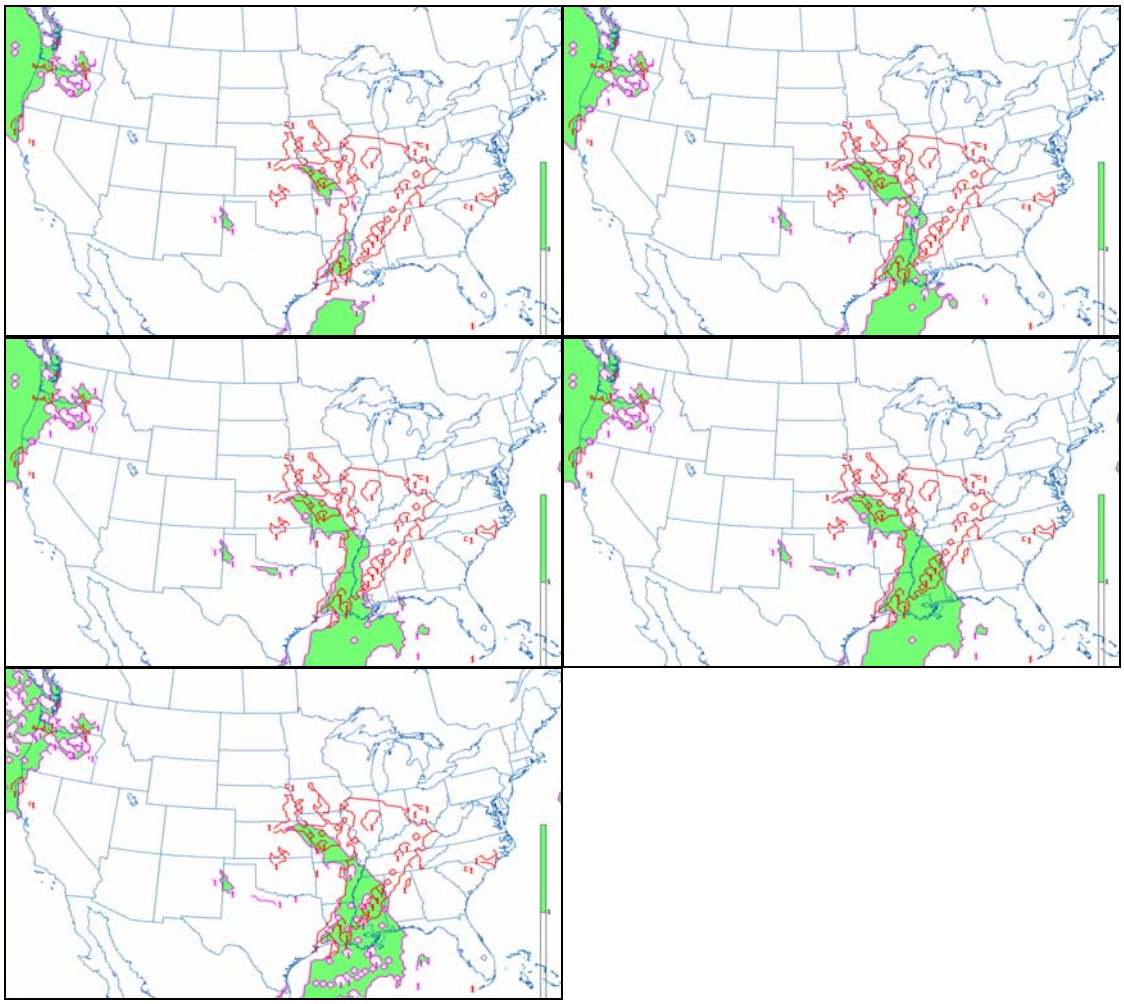

Figure 20. Configuration group F output for NAM 24 Feb 2007 0000Z run, 24-hour forecast (green shade, purple outline) and observed severe weather at 0000Z, 25 Feb 2007 (red outline). Configurations are ordered as follows: F1 (upper left), F2 (upper right), F3 (middle left), F4 (middle right), and F5 (lower left).

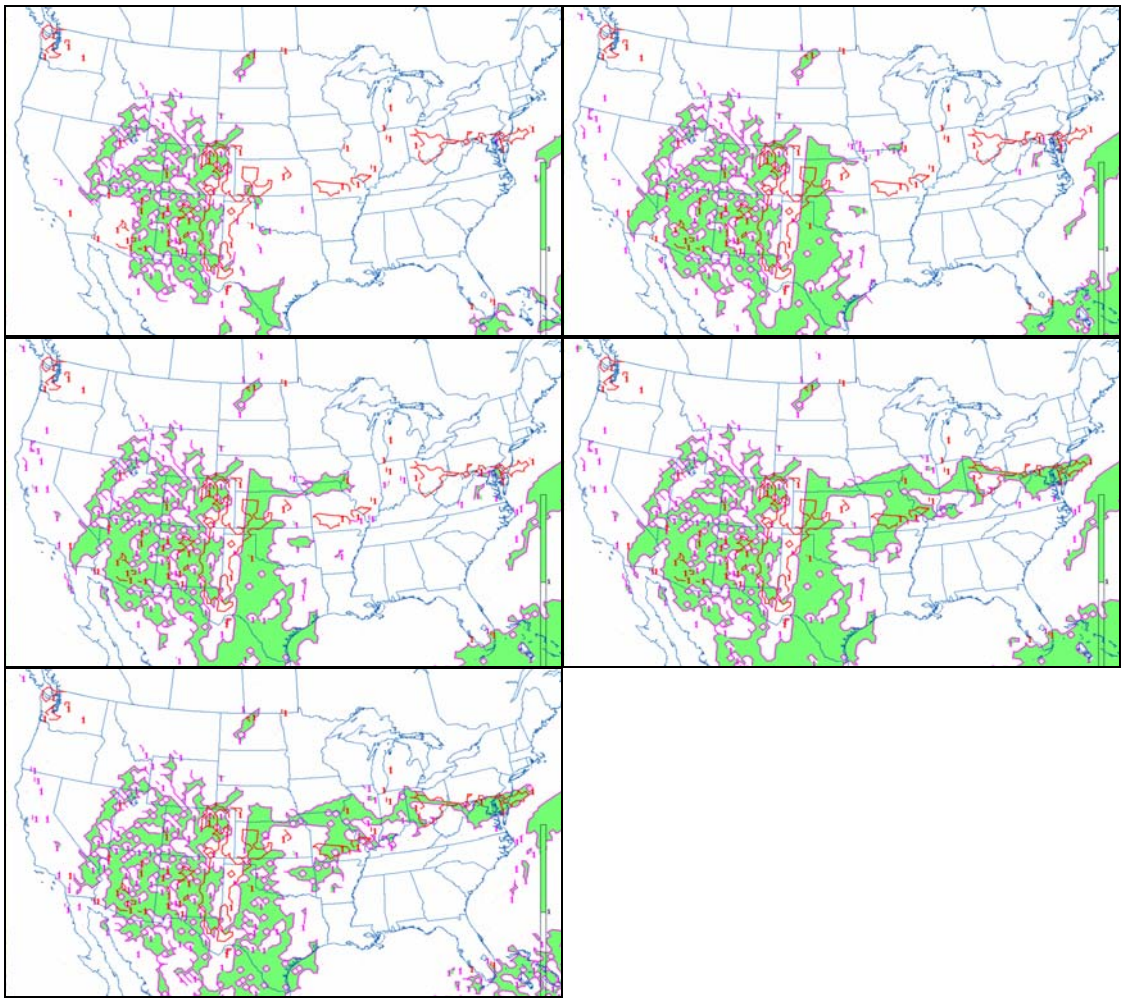

Figure 21. Configuration group F output for NAM 23 Mar 2007 1200Z run, 12-hour forecast (green shade, purple outline) and observed severe weather at 0000Z, 24 Mar 2007 (red outline). Configurations are ordered as follows: F1 (upper left), F2 (upper right), F3 (middle left), F4 (middle right), and F5 (lower left).

7. Discussion of Tier I Results

Based on the results from the Tier I experiments, configurations A6, A7, B5, C6, D4, E4, F4 and F5 were determined to have the highest performance ratings on the three cases. Therefore these eight configurations were retained for Tier II analysis across the entire dataset. It was also observed that configurations that use all or most of the original AFWA thresholds tended to have the poorest performance within each configuration group, which suggests that the thresholds defined by AFWA are set too high to capture a number of severe thunderstorm events to make the algorithm sufficiently reliable. Since AFWA originally developed and optimized the 3E Algorithm for use on the 45 km MM5 model, it is assumed the original settings perform optimally on MM5, although such is clearly not the case for the NAM model.

The 3E Algorithm had underforecast severe thunderstorm potential during the 1 February 2007 event along the Gulf Coast. This can be attributed to the occurrence of severe thunderstorms on both sides of a warm front situated from near New Orleans, Louisiana eastward to near Savannah, Georgia. Such severe weather cases present a particular challenge when warm, moist, and unstable air overrides a shallow cool layer of air near the surface. In such cases, the atmosphere may appear to not favor severe thunderstorms when relatively cool, stable conditions are present at the surface, but severe thunderstorms can still occur through elevated convection occurring in the warm, moist, and unstable air aloft.

Of the three cases, all configurations performed significantly better on the 24 February 2007 case than on the other two cases. This can be attributed to the fact that this particular case occurred from the southern Plains eastward into the Mississippi Valley, thus being a more “classic” severe thunderstorm event for which many of the severe thunderstorm indices were originally developed.

Finally, results from the 12-hour forecast on 23 March 2007 at 1200Z consistently showed very high false-alarm rates across all configuration groups. After further investigation it was determined the verification program was properly extracting and placing radar on the NAM grid following comparison with NWS archived radar images. It was further determined that a bad model run could be ruled out for the poor results, as

analysis of Skew-T soundings from Fort Worth (KFWD) and Corpus Christi (KCRP) both had Total-Totals and Lifted Index values that favored severe thunderstorms in addition to the presence of a weak baroclinic zone across this region. An analysis of radar imagery showed scattered convection with reflectivity values reaching the severe weather threshold of 40 dBZ between 23/1800Z and 23/2100Z. During this time period this convection was moving east-northeastward from western Texas toward central Texas. However, this activity began to dissipate after 23/2100Z and was largely absent by 24/0000Z, thus never reaching the favored severe thunderstorm region over central Texas. It is not immediately clear what caused severe thunderstorms to dissipate before reaching the area favored for continued severe convection.

Across all of the configurations and cases studied, the algorithm appeared to run at optimum performance when the threshold for CAPE was set at 2800 J kg^{-1}, the Total-Totals was set at 50, and the Lifted Index was set as high as -3.5. Additionally, analysis showed that CIN values of -90 J kg^{-1} or greater, were indicative of a sufficiently weak cap to favor the formation of severe thunderstorms. These findings suggest that previously established thresholds of 3000 J kg^{-1}, -50 J kg^{-1}, and 55 for CAPE, CIN, and Total-Totals, respectively, may be too high to capture a substantial portion of severe thunderstorm events. Likewise, analysis suggests that severe thunderstorms can occur at a much higher LI threshold of -3.5, vice the empirically-defined and widely accepted threshold of -5.0.

D. TIER II EXPERIMENT RESULTS

The eight configurations retained following the Tier I screening were run for all six severe thunderstorm events. It was hoped that the performance for these eight configurations would be as good or better than what was demonstrated on the three Tier I cases. However, Tier II analysis resulted in lower performance statistics for all eight configurations, although the trends for each show overall operational improvement over the default setting. This demonstrates the inherent challenges involved in using a synoptic to coarse mesoscale model in forecasting the occurrence of a severe thunderstorm event within a 32 km gridbox and a 3-hour time window. A side-by-side comparison of performance data from each configuration is presented in Figure 22.

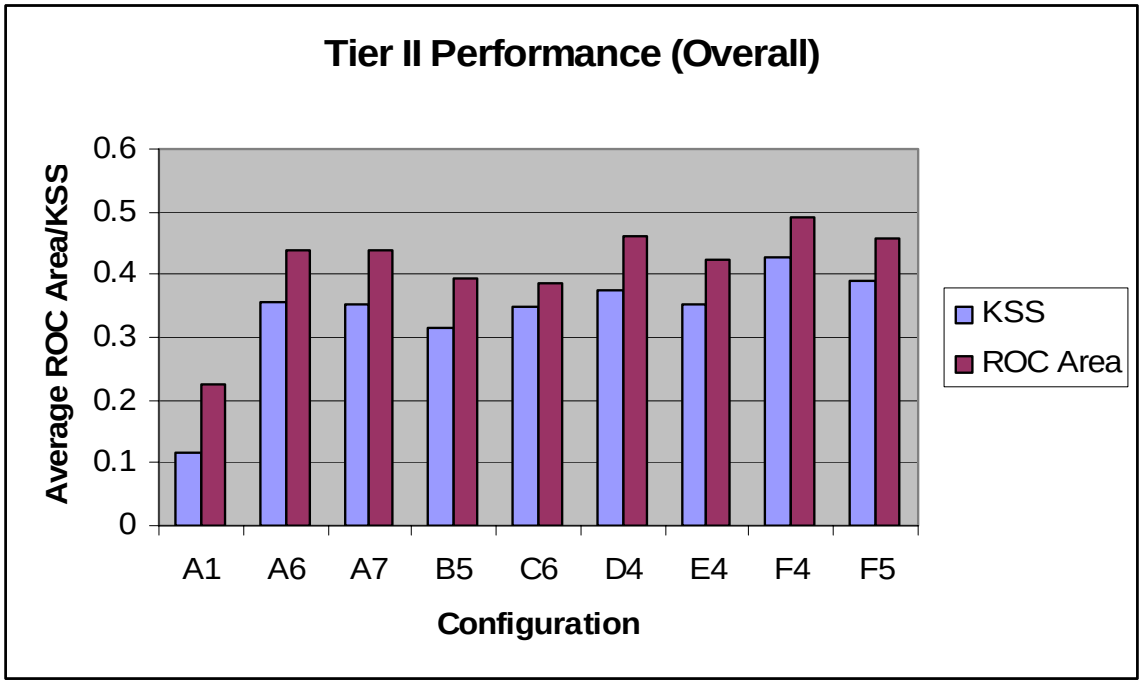

Figure 22. Overall performance of Tier II configurations.

The results in Figure 22 were obtained by running each algorithm configuration across all six severe thunderstorm cases, which yielded performance statistics for each model run and forecast time within the dataset. These statistics were then grouped and averaged according to forecast time to obtain an average KSS and ROC area for each forecast time. Finally, the overall performance statistics for a given configuration were compiled by summing the mean KSS and ROC values for each forecast time, then dividing by the number of forecast times (16; one for each 3-hour interval from forecast hour 03 to forecast hour 48). Based on the Tier II results, there was a net increase in both ROC area and KSS for all configurations over the AFWA default configuration A1. The largest increases occurred with configuration F4: its overall KSS of 0.426 represented an increase of 0.310 from the 0.116 KSS of configuration A1. Along the same lines the average ROC area of configuration F4 (0.4905), while still below the "no-skill" ROC area of 0.5, still reflects an increase of 0.2655 over the ROC area (0.2250) of configuration A1. Sample algorithm outputs using the default configuration for each severe thunderstorm event are presented in Figure 23. Note how the high proxy settings led to the forecast solutions being overconstrained, thus resulting in the severe thunderstorm threat region being underforecasted. Figures 24 through 29 are forecast

solutions of the experimental configurations for each case. The main issue for the default setting (A1) is it tended to underforecast the threat of severe thunderstorms, as depicted by Figure 23. This underforecasting suggested that the original thresholds for CAPE, Total-Totals Index, Lifted Index, and CIN were set too high to account for a large portion of severe thunderstorm events that occurred below these thresholds. The Tier II results were generally consistent with the findings from the Tier I experiments. Like in Tier I, data from Tier II indicated that decreasing the thresholds eased the constraint on the forecast solution, and thus the algorithm forecasted the threat of severe thunderstorms over a larger area that encompassed more of the region where such storms actually occurred. Modifications to the algorithm also helped it more accurately depict the synoptic-scale regions where severe thunderstorms can be expected. Some degree of overforecasting might actually be beneficial to potentially offset any model errors in the placement of synoptic-scale features throughout the forecast period.

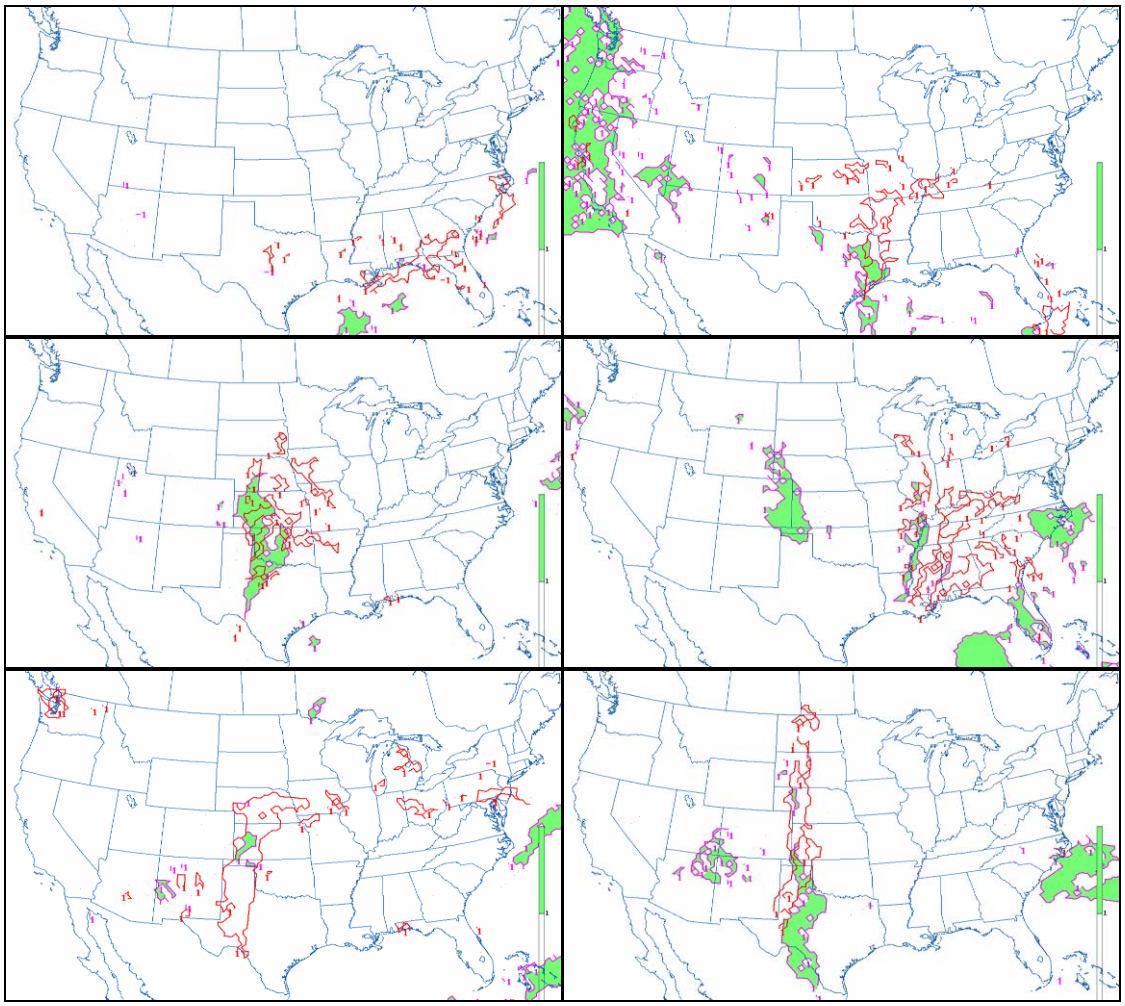

Figure 23. Algorithm output for configuration A1. Severe thunderstorms were forecasted by the algorithm in the green-shaded regions. The areas outlined in red indicate where severe thunderstorms occurred. The charts are samples from each case studied and are ordered as follows: 01 Feb 2007 12Z, 18-hour forecast (upper left); 12 Feb 2007 00Z, 21-hour forecast (upper right); 23 Feb 2007 12Z, 18-hour forecast (middle left); 28 Feb 2007 12Z, 30-hour forecast (middle right); 23 Mar 2007 12Z, 21-hour forecast (lower left); 28 Mar 2007 12Z, 15-hour forecast (lower right).

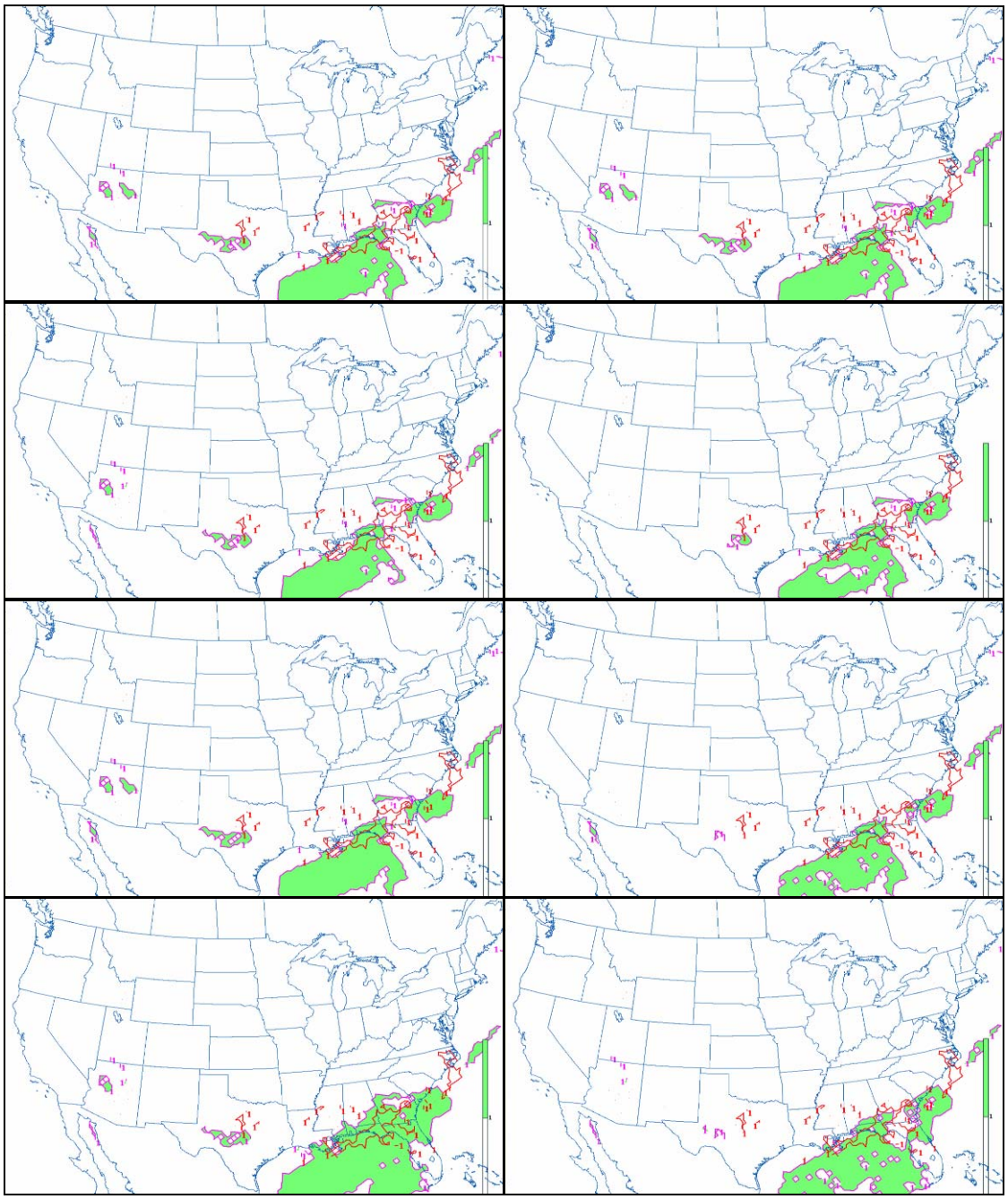

Figure 24. 3-Element Severe Thunderstorm Algorithm forecast using experimental configurations for NAM 12Z, 01 Feb 2007, 6-hour forecast. Severe thunderstorms were forecasted by the algorithm in the green-shaded regions. The areas outlined in red indicate where severe thunderstorms occurred. Configurations are ordered as follows: A6 (upper left), A7 (upper right), B5 (second row left), C6 (second row right), D4 (third row left), E4, third row right), F4 (lower left), and F5 (lower right).

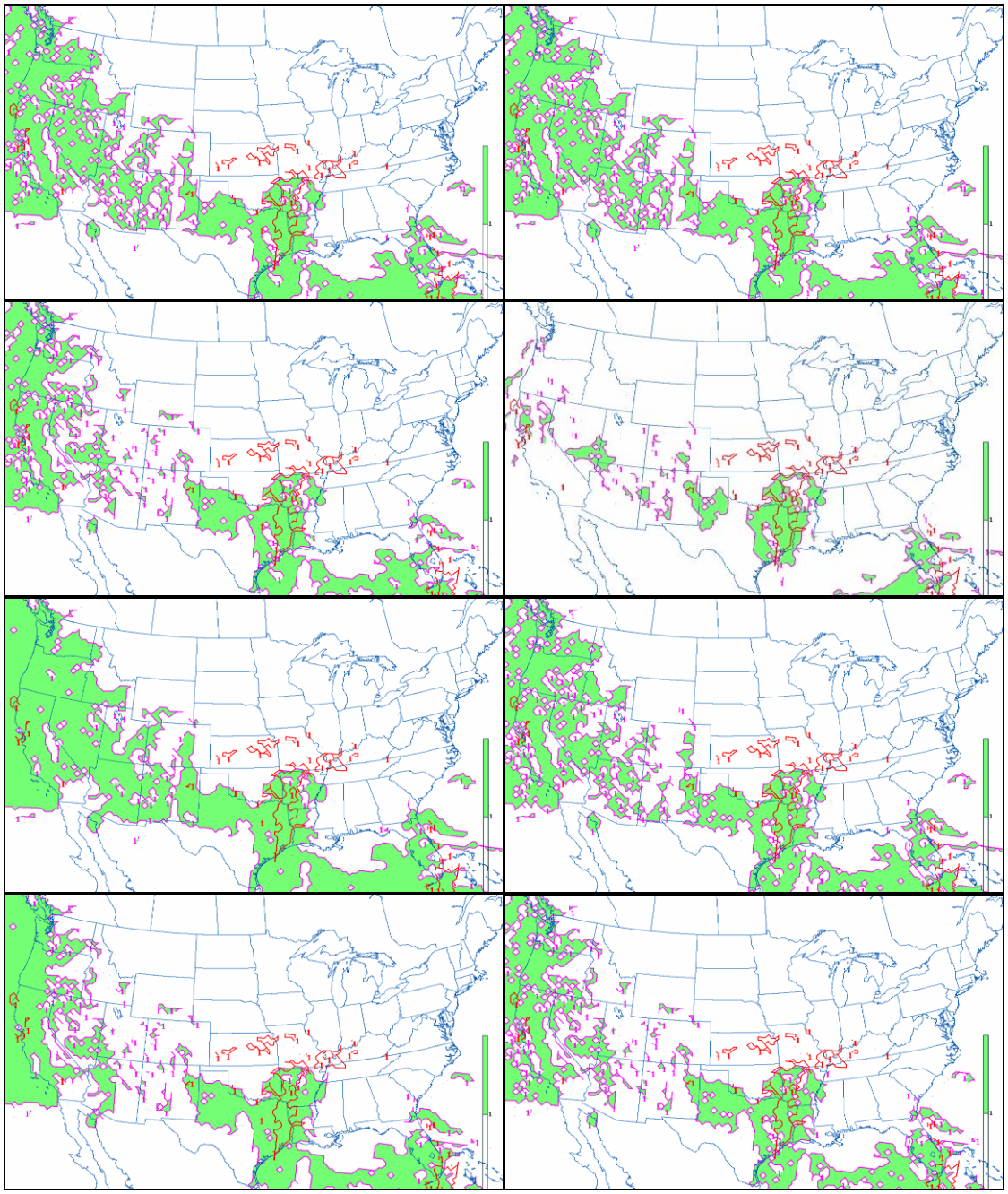

Figure 25. 3-Element Severe Thunderstorm Algorithm forecast using experimental configurations for NAM 00Z, 12 Feb 2007, 21-hour forecast. Severe thunderstorms were forecasted by the algorithm in the green-shaded regions. The areas outlined in red indicate where severe thunderstorms occurred. Configurations are ordered as follows: A6 (upper left), A7 (upper right), B5 (second row left), C6 (second row right), D4 (third row left), E4, third row right), F4 (lower left), and F5 (lower right).

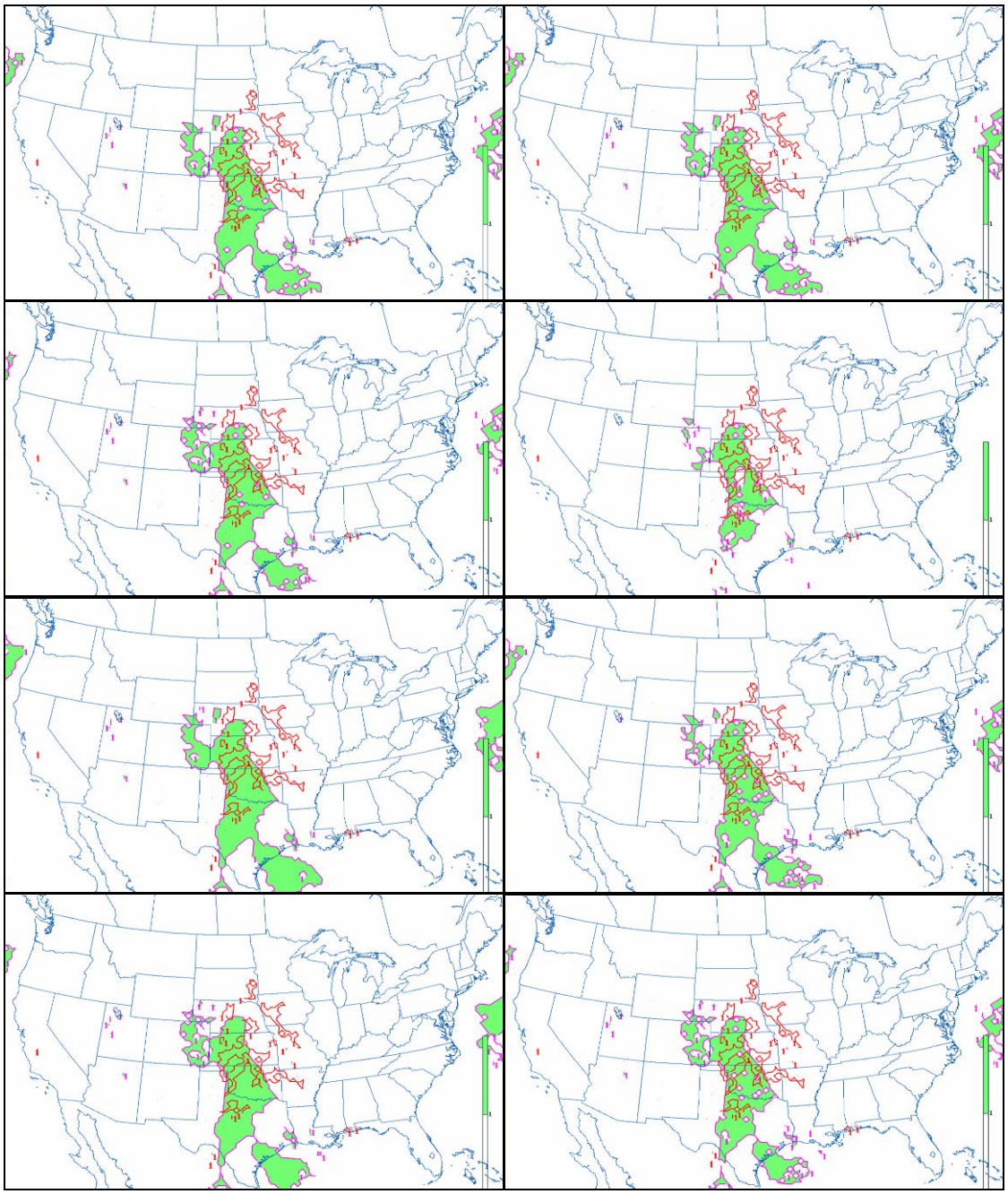

Figure 26. 3-Element Severe Thunderstorm Algorithm forecast using experimental configurations for NAM 12Z, 23 Feb 2007, 18-hour forecast. Severe thunderstorms were forecasted by the algorithm in the green-shaded regions. The areas outlined in red indicate where severe thunderstorms occurred. Configurations are ordered as follows: A6 (upper left), A7 (upper right), B5 (second row left), C6 (second row right), D4 (third row left), E4, third row right), F4 (lower left), and F5 (lower right).

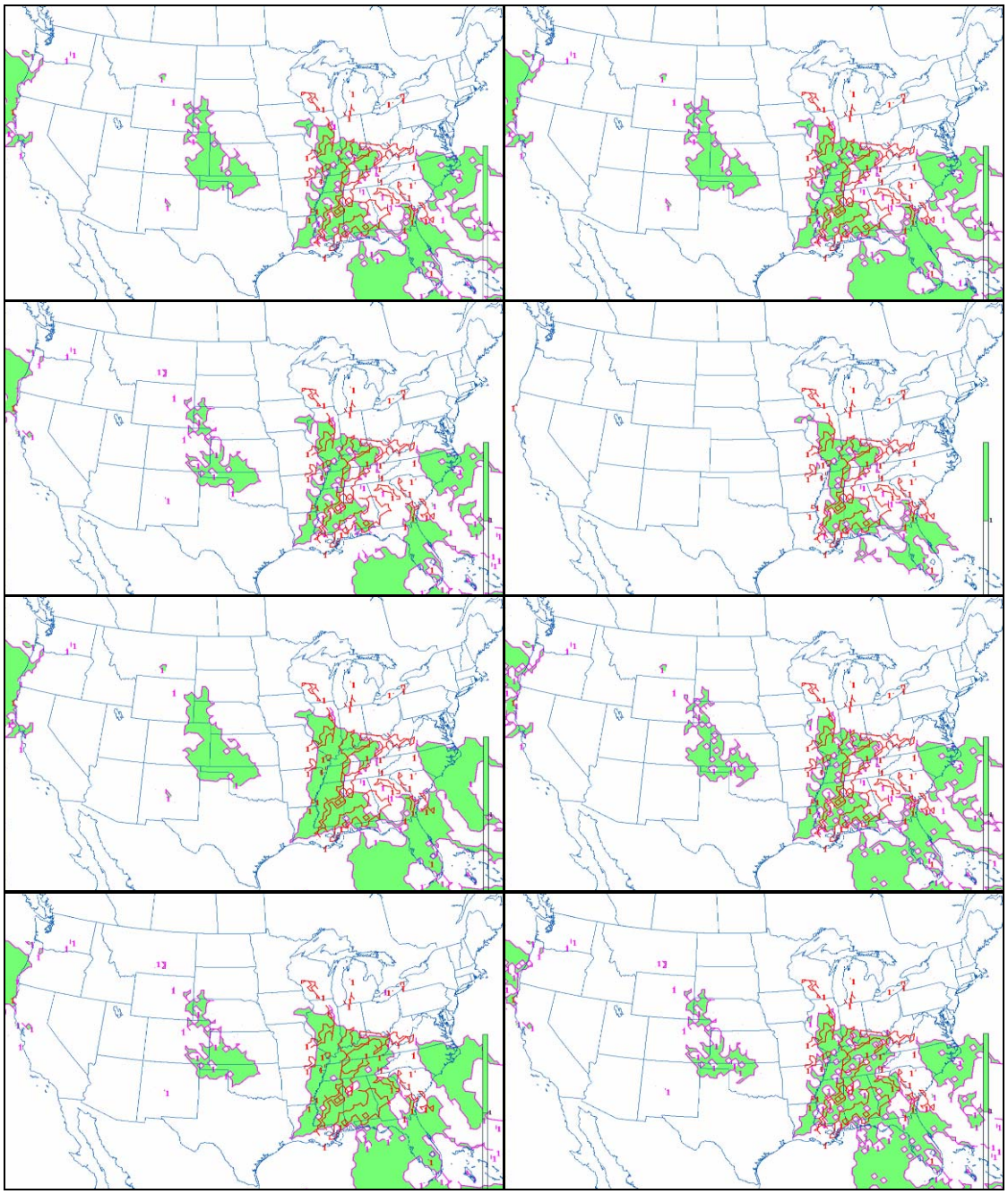

Figure 27. 3-Element Severe Thunderstorm Algorithm forecast using experimental configurations for NAM 12Z, 28 Feb 2007, 30-hour forecast. Severe thunderstorms were forecasted by the algorithm in the green-shaded regions. The areas outlined in red indicate where severe thunderstorms occurred. Configurations are ordered as follows: A6 (upper left), A7 (upper right), B5 (second row left), C6 (second row right), D4 (third row left), E4, third row right), F4 (lower left), and F5 (lower right).

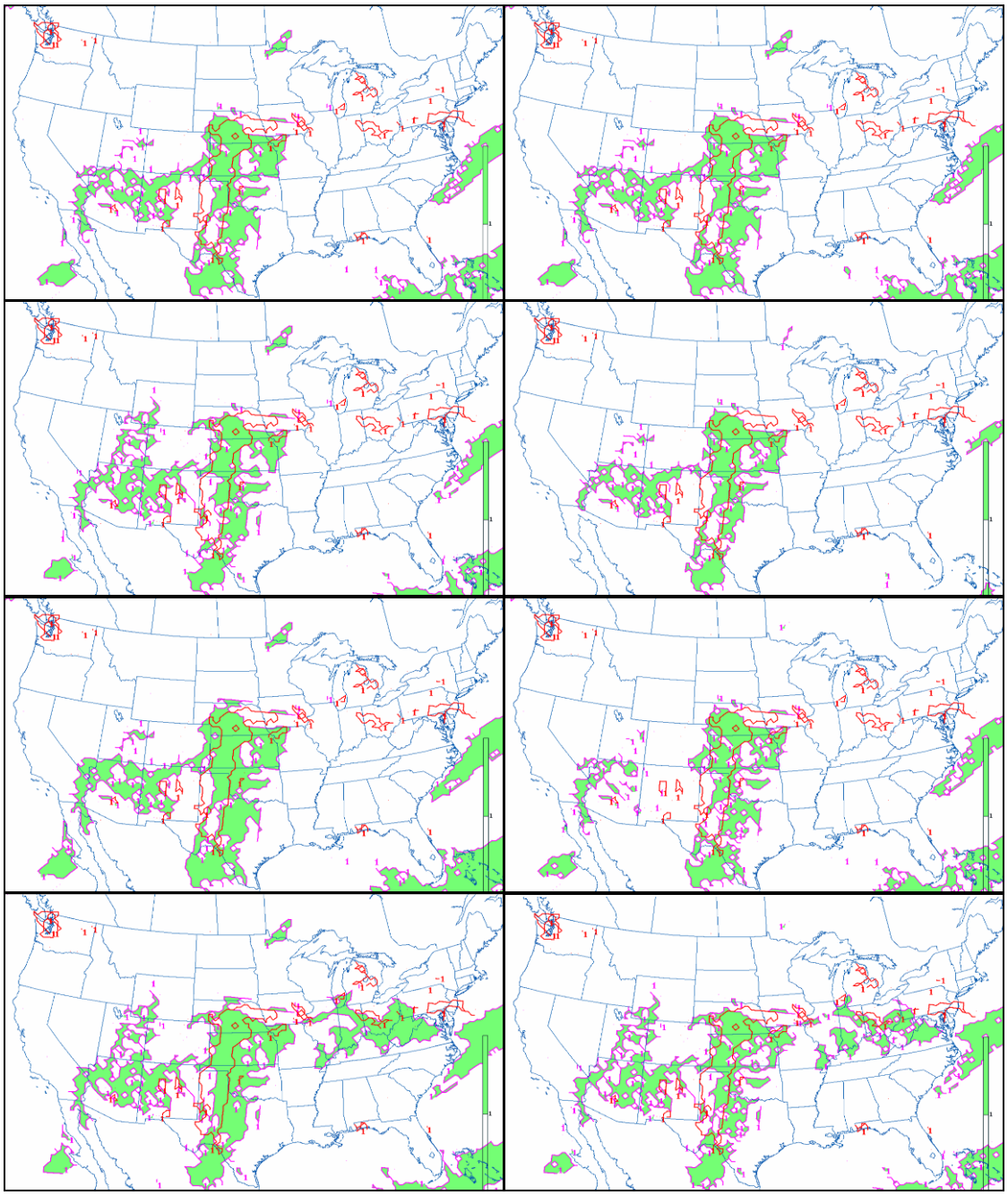

Figure 28. 3-Element Severe Thunderstorm Algorithm forecast using experimental configurations for NAM 12Z, 23 Mar 2007, 21-hour forecast. Severe thunderstorms were forecasted by the algorithm in the green-shaded regions. The areas outlined in red indicate where severe thunderstorms occurred. Configurations are ordered as follows: A6 (upper left), A7 (upper right), B5 (second row left), C6 (second row right), D4 (third row left), E4, third row right), F4 (lower left), and F5 (lower right).

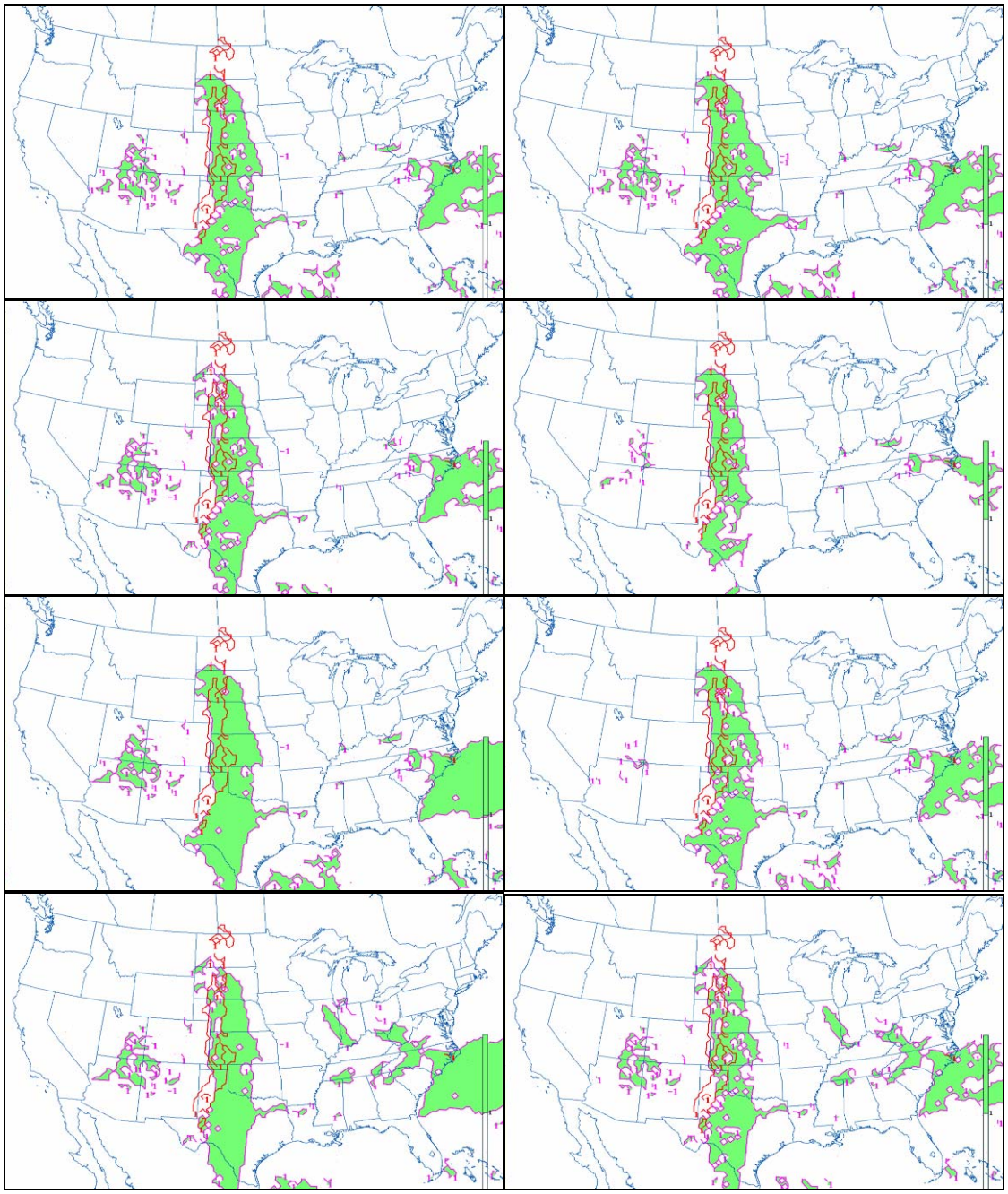

Figure 29. 3-Element Severe Thunderstorm Algorithm forecast using experimental configurations for NAM 12Z, 28 Mar 2007, 15-hour forecast. Severe thunderstorms were forecasted by the algorithm in the green-shaded regions. The areas outlined in red indicate where severe thunderstorms occurred. Configurations are ordered as follows: A6 (upper left), A7 (upper right), B5 (second row left), C6 (second row right), D4 (third row left), E4, third row right), F4 (lower left), and F5 (lower right).

V. SUMMARY AND CONCLUSION

Analysis on the AFWA 3-Element Severe Weather Forecast Algorithm was performed with 30 configurations on seven severe thunderstorm cases using a two-tiered experimental approach. All 30 configurations were tested on three random model forecasts during the Tier I experiment. Following Tier I analysis, the original set of 30 configurations was reduced to eight for analysis across the entire dataset during the Tier II experiment. Based on the results, the AFWA default setting performed with an overall ROC area of 0.2250 and KSS of 0.116. Configuration F4 represented the greatest improvement over the default setting, and is therefore the optimal setting for the 32 km NAM model. The ROC area increased by 0.2655 and the KSS increased by 0.310 when the algorithm was run using configuration F4. While the remaining configurations also improved the algorithm's performance, those were less than the improvements observed with the optimal setting.

It is further concluded that amount of improvement in the 3E Algorithm's performance is constrained by the performance of the computer model on which it is run. Along those lines, any errors in the model output will propagate into the 3E Algorithm and adversely impact the algorithm's performance. This is not a flaw of the 3E Algorithm per se, but it is a factor that must be considered when evaluating the performance of this and any other algorithms or products that rely on model output data.

Another point of concern was the fact that most of the severe thunderstorm indices integrated into the 3E Algorithm were developed through extensive research of severe thunderstorms that occurred mainly in the southern Plains. As such the widely accepted thresholds for these indices were established and optimized for this specific region. The same indices and/or thresholds are not necessarily optimal for accurately forecasting severe thunderstorms in other regions where somewhat different conditions may be required for the initiation of severe thunderstorms. A particular challenge was forecasting severe thunderstorms that initiate in a warm, unstable air mass overlying a cooler, more stable layer near the surface. Similarly, modifications to the indices were

necessary to account for high-elevation regions where the lower bounds used to compute index values may be at or below ground elevation.

It is also important to consider that this study attempted to generate a product that forecasted the potential for severe thunderstorms, and verification was constrained to a 32 km gridbox with a 3-hour time window. In reality, severe thunderstorms occur on space and time scales that are too small to be accurately resolved by the 32 km NAM in its present form. Additionally, this study proved there were significant issues in the development and implementation of an algorithm with the intention of global application for forecasting severe thunderstorms. While the three elements of instability, a weak cap, and dynamic forcing generally favor the formation of severe thunderstorms, the minimum required amounts of each element varied depending on geographic area, season, and even time of day. The 3-Element Severe Weather Forecast Algorithm is not well-suited at this time for accurately pinpointing the exact time and place a severe thunderstorm will form, however it did effectively indicate the general area where the potential for severe weather exists. Therefore, it should be used as a guidance tool to alert the meteorologist to areas that demand more scrutiny in terms of assessing the threat for severe weather and appropriately monitoring such regions for rapidly changing conditions. This experiment was performed on six severe thunderstorm events, which is a limited dataset. It is possible that running the default and optimal configurations on a much more robust dataset containing several dozen severe thunderstorm events spanning multiple years, could yield much better statistical results above those gathered in this study with the focus on synoptic accuracy rather than attempting to pinpoint the exact time and location of severe thunderstorm occurrence.

VI. FUTURE RESEARCH

When the 3-Element Severe Weather Forecast Algorithm indicates favorable conditions for severe thunderstorms, the meteorologist would naturally think, “What type of severe weather can I expect?” Future studies should attempt to adapt the 3E Algorithm to determine the type of severe weather (i.e., tornadoes, damaging winds, or large hail) most likely to occur in the favored development region. Since the type of severe convection is dependent on the relationship between vertical shear and buoyancy (Harnack, et. al. 1997, and Rotunno 1984), future research should consider the relationship between shear and buoyancy, and establishing thresholds for single-cell, multicellular lines and mesoscale convective systems, and tornadic supercells.

Finally, the 3E Algorithm should be verified and optimized to run on multiple model platforms and ensembles. There are two potential options for integrating the algorithm into ensemble forecasting. The first option is to run multiple algorithm configurations on a single model, then using the spread of forecast solutions to identify areas where the different configurations are in agreement. The other method is to run the 3E Algorithm on multiple model platforms in the optimal settings for each model, then analyze the different model solutions to identify the regions with the greatest severe weather threat. Finally, the number of possible configurations for the 3E Algorithm is virtually endless. As such, the algorithm should be experimented using additional configurations not developed or included in this study to determine their impacts on algorithm performance.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A: FORMULAS FOR COMPUTING ELEMENT PROXIES

The following equations are used to compute some of the proxies used by the 3E Algorithm. Formulas for CAPE, CIN, and Lid Strength Index are from Andrew Revering's List of Meteorological Formulas web page, which can be accessed at http://www.aprweather.com/pages/calc.htm. Equations for Total-Totals Index, K-Index, and Lifted Index are from Miller (1972).

Total-Totals Index: $TT = T_{850mb} + T_{d,850mb} - 2(T_{500mb})$

Elevated Total-Totals Index: $ETT = T_{700mb} + T_{d,700mb} - 2(T_{500mb})$

K-Index: $KI = (T_{850mb} - T_{500mb}) + T_{d,850mb} - (T_{700mb} - T_{d,700mb})$

850mb Thermal Advection: $\left(\vec{V}\square\vec{\nabla}T\right)_{850mb} = \left(u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y}\right)_{850mb}$

700mb Thermal Advection $\left(\vec{V}\square\vec{\nabla}T\right)_{700mb} = \left(u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y}\right)_{700mb}$

Surface Based CAPE and Convective Inhibition are computed by lifting an air parcel dry adiabatically from the surface to the lifting condensation level (LCL). The parcel is then lifted moist adiabatically from the LCL to the equilibrium level. For layers where the parcel temperature is greater than the environmental temperature, CAPE is computed based on the following equation:

$$CAPE = g\int_{LFC}^{EL}\left(\frac{T_p - T}{T}\right)dz\,;\; T_p > T\,.$$

The above equation is computed by the 3E Algorithm using the iterative approximation: $CAPE = g\sum_{LFC}^{EL}\left(\frac{T_p - T}{T}\right)_i (z_i - z_{i+25mb})$, where g is the acceleration due to gravity (9.8 ms^{-2}); LFC is the level of free convection; EL is the equilibrium level; T_p is the parcel temperature; T is the environmental temperature; (z_i-z_{i+25mb}) is the thickness of the

iterative layer in meters. Convective inhibition is computed in the same manner as CAPE, but for layers where the parcel temperature is lower than the environmental temperature:

$$CIN = g \int_{LFC}^{EL} \left(\frac{T_p - T}{T} \right) dz \; ; \; T_p < T \; .$$

Convective inhibition is approximated using the same iterative method used for approximating CAPE, keeping in mind this applies for layers where the parcel temperature is colder than the environment: $CIN = g \sum_{LFC}^{EL} \left(\frac{T_p - T}{T} \right)_i (z_i - z_{i+25mb}) \; .$

Lifted Index is obtained by lifting an air parcel dry adiabatically from the surface to the LCL, then continue lifting the parcel moist adiabatically from the LCL to 500mb. The Lifted Index is computed by subtracting the parcel temperature from the environmental temperature.

APPENDIX B: AIRCRAFT CATEGORIES

Air Force aircraft are categorized according to their sensitivity to turbulence, which depends on the weight, wing surface area, and wind sweep angle of the aircraft (AFWA/DNT 1998). This list is current as of the 15 July 1998 publication of AFWA TR-98/002, and is not all-inclusive.

Category	**Aircraft**
I	OH-58, UH-1, AH-1
II	C-141, C-9, RAH-66, C-12, C-21, F-106, C-20, C-5A, E-4A, F-15, AH-64, B-52, C-130, C-17, F-117, F-16, KC-135, C-23, CH-47, U-2, OV-1, CH-3, UH-60, CH-53, CH-54, VC-137, T-38
III	OV-10, KC-10, T-37, A-10
IV	A-7, F-4, B-1, F-111

Table 31. United States Air Force aircraft categories.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C: FORMULAS FOR COMPUTING STATISTICS

Using the contingency diagram from Fowle and Roebber (2003), the total number of hits (A), "false alarms" (B), misses (C), and correct non-events (D) were entered into the following equations (Fowle and Roebber 2003; Lee and Passner 1993) to compute performance statistics for a given model run and forecast forecast time.

Probability of Detection: $$POD = \frac{A}{(A-C)}$$

Miss Rate: $$MR = 1 - POD$$

False Alarm Rate: $$FAR = \frac{B}{(A-C)}$$

Bias $$BIAS = \frac{(A+B)}{(A-C)}$$

Threat Score: $$TS = \frac{A}{(A+B+C)}$$

Equitable Threat Score: $$ETS = \frac{(A-E)}{(A+B+C-E)}$$

Correct Nonoccurrence: $$CN = \frac{D}{C+D}$$

Kuiper Skill Score: $$KSS = \frac{(AD-BC)}{(A+C)(B-D)}$$

Chance Event Correction: $$E = \frac{(A+C)^2}{(A+B+C+D)}$$

Relative Operating Characteristic (ROC) is computed by first plotting the Probability of Detection (POD) versus False Alarm Rate (FAR) on the ROC Diagram

(Figure 30). In the ROC Diagram, FAR is plotted in the x-direction and POD is plotted in the y-direction. The following equation is used to compute ROC area for a single point on the ROC diagram.

ROC Area:

$$A_{ROC} = 0.50 \pm \frac{\sqrt{2}}{4}\sqrt{(POD - FAR)^2 + (POD - FAR)^2}$$

When the POD is greater than the FAR, add the second term to 0.50, and subtract the second term from 0.50 when the POD is less than the FAR. A ROC area of 0.50 is regarded as a "no-skill" forecast, while a ROC area above 0.50 represents an improvement in forecast skill. A ROC area of 1.0 is considered a perfect forecast.

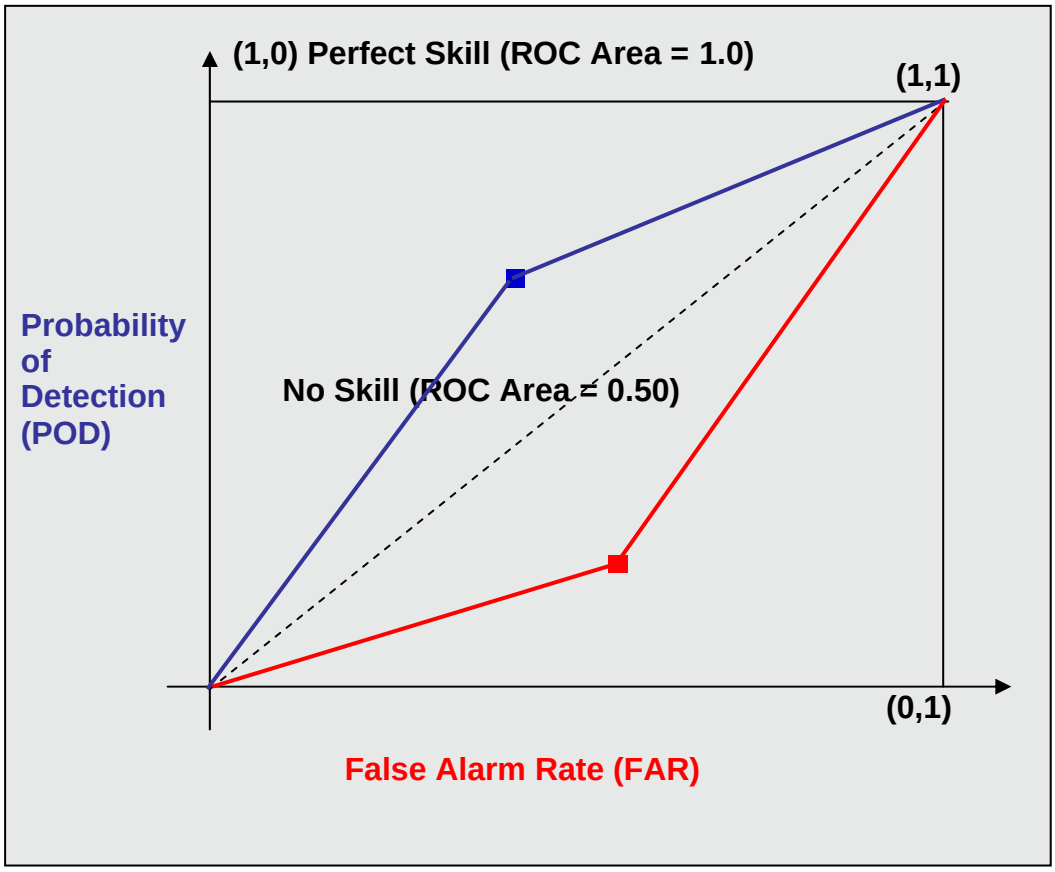

Figure 30. Relative Operating Characteristic (ROC) Diagram used to compute ROC area. The point intersected by the blue line represents a forecast that surpasses the "no-skill" forecast. The point bisected by the red line indicates a forecast that does not meet the "no-skill" ROC area.

LIST OF REFERENCES

AFWA/DNT, 1998: TN-98/002, Meteorological Techniques, Headquarters Air Force Weather Agency, Offutt Air Force Base, Nebraska, 242 pp.

Air Force Weather Agency Meteorological Models Branch, "MM5 Model Physics" https://weather.afwa.af.mil/HOST_HOME/DNXM/ABOUTMM5/toc/physics/index.html, Retrieved February 2008.

Fowle, Michael A., and Paul J. Roebber, 2003: Short Range (0-48h) Numerical Prediction of Convective Occurrence, Mode, and Location, *Wea. Forecasting*, **18**, 782-793.

Harnack, Robert P., Donald T. Jensen, Joseph R. Cermak III, 1997: Investigation of Upper-Air Conditions Occurring with Warm Season Severe Wind Events in Utah, *Wea. Forecasting*, **12**, 282-293.

Johnson, J.T., Pamela L. MacKeen, Arthur Witt, E. DeWayne Mitchell, Gregory J. Stumpf, Michael D. Eilts, and Kevin W. Thomas, 1997: The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm, *Wea. Forecasting*, **13**, 263-276.

Keller, David L., 2004: "An Automated 3-Element Algorithm for Forecasting Severe Weather Using AFWA MM5 Model Output Data." 22nd Conf. on Severe Local Storms. Hyannis, MA, Amer. Meteor. Soc., CD-ROM P11.5.

Lee, Robert R., and Jeffrey E. Passner, 1993: The Development and Verification of TIPS: An Expert System to Forecast Thunderstorm Development, *Wea. Forecasting*, **8**, 271-280.

Miller, R.C., 1972: AWS Technical Report 200 (Rev.), Headquarters, Air Weather Service, Scott AFB, IL, 106 pp.

Rasmussen, Erik N., and David O. Blanchard, 1997: A Baseline Climatology of Sounding-Derived Supercell and Tornado Forecast Parameters, *Wea. Forecasting*, **13**, 1148-1164.

Revering, Andrew, "Andrew Revering's List of Meteorological Formulas," http://www.aprweather.com/pages/calc.htm, Retrieved August 2007.

Rotunno, Richard, 1984: Tornadoes and Tornadogenesis, *Mesoscale Meteorology and Forecasting*, 415-436.

Storm Prediction Center, http://www.spc.noaa.gov. Retrieved August 2007.

Thompson, Richard L., Corey M. Mead, and Roger Edwards, 2005: Effective Storm-Relative Helicity and Bulk Shear in Supercell Thunderstorm Environments, *Wea. Forecasting*, **22**, 102-115.

USAF/XOWP, 2004: Air Force Manual 15-129, Air and Space Operations – Processes and Procedures, Headquarters United States Air Force, Washington, DC, 161 pp.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Air Force Weather Agency
   Offutt AFB, Nebraska